

On Implementation Aspects of Decode and Forward and Compress and Forward Relay Protocols

Dissertation zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.) der Fakultät für Informatik und
Elektrotechnik der Universität Rostock

vorgelegt von

Daniel Kern, geboren am 18. Februar 1987 in Rostock

Jahr der Einreichung: 2018

Jahr der Verteidigung: 2018

Gutachter:

Prof. Dr.-Ing. habil. Volker Kühn, Universität Rostock, Institut für Nachrichtentechnik

Prof. Dr.-Ing. habil. Tobias Weber, Universität Rostock, Institut für Nachrichtentechnik

Prof. Dr.-Ing. Peter A. Höher, Christian-Albrechts-Universität zu Kiel

https://doi.org/10.18453/rosdok_id00002380

PhD Thesis

Faculty of Computer Science and Electrical Engineering

Danksagung

An erster Stelle gilt mein Dank meinem Doktorvater Prof. Volker Kühn für seine wissenschaftliche und methodische Unterstützung sowie Motivation und Hifestellung während meiner gesamten Zeit am Institut für Nachrichtentechnik der Universität Rostock. Ich bin dankbar, dass er mir die Möglichkeit gegeben hat, fünf unvergessliche Jahre in einem sowohl familiären als auch sehr professionellen Umfeld zu wachsen. Weiterhin danke ich meinen Kollegen, die mich bei meiner Arbeit stets mit zielführenden Diskussionen und Ratschlägen unterstützt haben. Besonderers danken möchte ich Dr.-Ing. Henryk Richter, Dr.-Ing. Sebastian Vorköper, Peter Bartmann und Jan Lewandowsky für ihre Anmerkungen und Korrekturen zur Erstellung der finalen Version meiner Dissertation. Besonderer Dank gilt auch meinen Eltern und meiner Lebensgefährtin Katharina, die mir auf dem Weg zur fertigen Arbeit Raum und Zeit einräumten und mir motivierend zur Seite standen.

Zusammenfassung

Diese Arbeit untersucht die Relay-Protokolle Decode-and-Forward (DF) und Compress-and-Forward (CF) mit dem Fokus auf einer praktischen Umsetzung. Es werden zum Einen in der Theorie vernachlässigte Störeinflüsse wie Kanal- und Phasenschätzfehler betrachtet. Zum Anderen steht die Frage der Realisierung einzelner Signalverarbeitungsschritte im Vordergrund. Im Speziellen ist hierbei die Quantisierung für CF von großer Bedeutung, weil diese in der Informationstheorie lediglich durch Quantisierungsrauschen modelliert wird. Zur Auswertung der Leistungsfähigkeit der Protokolle werden die jeweiligen erzielbaren Raten entweder numerisch mit Hilfe des Max-Flow Min-Cut (MFC) Theorems bestimmt oder durch Simulation eines Bit Interleaved Coded Modulation (BICM) Systems, welches eine dem Stand der Technik entsprechende Turbocodierung sowie Quadraturamplitudenmodulation verwendet.

Für das DF-Protokoll spielt neben der Superpositionsmodulation auch verteiltes Beamforming eine bedeutende Rolle, weshalb der Einfluss von Phasenfehlern mit Hilfe eines statistischen Modells analysiert wird. Sowohl numerische als auch simulative Ergebnisse bestätigen, dass die Leistungsfähigkeit in Abhängigkeit der Varianz des Phasenfehlers abnimmt. Des Weiteren wurde Superpositionsmodulation verwendet, um das klassische informationstheoretische Prinzip der Superpositionsmodulation in einen Mehrstufencode zu verallgemeinern, wodurch ein "Shaping"-Gewinn inhärent erzielt wird.

Da bezüglich CF gleichförmige und auch die Entropie maximierende Quantisierer aufgrund von Empfängerrauschen ungeeignet sind, wurde die sogenannte Information Bottleneck (IB) Methode verwendet, um möglichst optimale Quantisierer zu entwerfen. Da die Komplexität eines IB-Algorithmus' unter anderem vom Kanalmodell abhängt, kann es sein, dass der benötigte Speicher oder die Laufzeit die Kapazitäten aktueller Hardware weit übersteigt. In diesem Fall ist es notwendig das Problem mit Hilfe von Näherungen zu vereinfachen. Zu diesem Zweck werden IB-Graphen entwickelt, um verschiedene Ansätze solcher Vereinfachungen herzuleiten: Als Übertragungsverfahren wird Orthogonal Frequency Division Multiplexing (OFDM) angenommen, wobei entsprechende Subträger als Rayleigh-Fading-Kanäle modelliert werden. Abschließend werden auch Kanalschätzfehler betrachtet, wobei Kanalschätzung und Entzerrung entweder implizit durch den Quantisierer erfolgen oder explizit vor der Quantisierung durchgeführt werden.

Abstract

In this work, the common relay protocols Decode-and-Forward (DF) and Compress-and-Forward (CF) are investigated from a practical point of view: This involves on the one hand the impact of imperfections like channel and carrier phase estimation errors and on the other hand, the question of how to implement relay protocol specific signal processing like quantization for CF which is modeled in information theory simply by additive quantizer noise. To evaluate the performance, achievable rates are determined either numerically with the help of the Max-Flow Min-Cut (MFMC) theorem or by link level simulations whereby a Bit Interleaved Coded Modulation (BICM) scheme is set up with state-of-the-art turbo coding and Quadrature Amplitude Modulation (QAM).

As distributed beamforming plays a major role for DF, the influence of phase deviations is addressed by means of proper statistical modeling. Both numerical and simulation results confirm a performance degradation depending on the phase error variance. Moreover, superposition modulation is used to create a multilevel code which generalizes the original information theoretic concept of superposition coding to exploit a shaping gain.

For CF, uniform or Maximum Output Entropy (MOE) quantizers are usually not appropriate due to the irrelevant noise in the relay's receive signal. Hence, the Information Bottleneck (IB) method is used to determine a suitable quantization scheme which is optimal except for numerical limitations. Corresponding quantizer mappings are determined and integrated into the simulation environment. Depending on the channel model, storage and also run time complexity of an IB algorithm may grow infeasibly large. In such cases, approximations are necessary to find a suitable quantizer. Therefore, Information Bottleneck Graphs (IBGs) are used to derive and evaluate different potentially suboptimal but feasible quantizer designs for an Orthogonal Frequency Division Multiplexing (OFDM) system with Rayleigh fading subcarriers. Finally, imperfect channel knowledge is considered whereby channel estimation and equalization is either implicitly done during quantization or explicitly before it.

Thesen

1. Decode and Forward approaches capacity for scenarios where the relay is considerably closer to the source than to the destination.
2. The beamforming gain of Decode and Forward is greater the closer the relay is placed towards the source.
3. Imperfect carrier phase synchronization degrades the beamforming performance but makes it not obsolete until phase errors are severe.
4. The beamforming performance decreases with growing variance of a random phase error.
5. Superposition modulation is well suited to implement the cooperative transmission strategy of source and relay as multilevel code.
6. Compress and Forward outperforms Decode and Forward when the source-relay link becomes the bottleneck of the system.
7. Maximum output entropy quantization is inappropriate for compression at the relay because of additive receiver noise.
8. The information bottleneck method delivers optimal quantization schemes for digital communication.
9. Storage complexity becomes an issue for information bottleneck algorithms when the cardinality or dimensionality of the problem is very high as for frequency selective channels.
10. Information bottleneck graphs are suitable to visualize and simplify information bottleneck problems to design quantization schemes for multi-carrier transmission.

Contents

Acronyms	XIII
Symbols	XVII
List of Figures	XIX
List of Tables	XXIII
1 Introduction	1
1.1 Motivation and Objective	1
1.2 Thesis Outline	2
1.3 Notation	3
2 Fundamentals - Point to Point Transmission	5
2.1 Channel Model	5
2.1.1 AWGN Channel with Path Loss	6
2.1.2 Rayleigh Fading	6
2.2 Channel Capacity	7
2.3 Modulation and Coding Schemes	7
2.3.1 Quadrature Amplitude Modulation	8
2.3.2 Turbo Coding	9
2.3.3 Multilevel Coding	12
2.3.4 Superposition Modulation	14
2.3.5 Bit Interleaved Coded Modulation	21
2.4 Simulation Setup and Achievable Rates	22
2.5 Orthogonal Frequency Division Multiplexing	22
2.5.1 OFDM - System Setup	23
2.5.2 Channel Estimation and Equalization	24
2.6 Discussion	25
3 Classical Relay Channel - Preliminaries	27

3.1	System Setup – 3-Node Relay Channel	27
3.2	Capacity Upper Bound	28
3.2.1	Full-Duplex Relays	28
3.2.2	Half-Duplex Relays	29
3.3	Relaying Strategies and Achievable Rates	29
3.3.1	Amplify and Forward	30
3.3.2	Decode and Forward	31
3.3.3	Compress and Forward	33
3.4	The Gaussian Relay Channel	35
3.4.1	Capacity Upper Bound	35
3.4.2	Amplify and Forward	35
3.4.3	Decode and Forward	36
3.4.4	Compress and Forward	36
3.5	Achievable Rates for the Gaussian Relay Channel	37
3.5.1	Individual Per Node Power Constraint	38
3.5.2	Sum Energy Constraint	40
3.5.3	Individual Per Node Energy Constraint	45
3.6	Discussion	45
4	Superposition Multilevel Coding for Decode and Forward	47
4.1	General Distributed Superposition Multilevel Code	47
4.2	Distributed Superposition Multilevel Code with Parallel Detection	50
4.3	Achievable Rates for Superposition Multilevel Coding	53
4.4	Discussion	54
5	Decode and Forward for Perfect and Imperfect Carrier Synchronization	55
5.1	Distributed Beamforming with Phase Errors	55
5.2	Random Phase Error Model	56
5.3	Max Flow Min Cut Theorem with Phase Errors	57
5.4	Rate Allocation for Practical System with Phase Errors	60
5.4.1	Orthogonal Channel Access	60
5.4.2	Non-Orthogonal Channel Access	61
5.5	Achievable Rates for Decode and Forward with Phase Errors	63
5.5.1	Results for Max Flow Min Cut Theorem	63
5.5.2	Simulation Results for Bit Interleaved Coded Modulation Scheme	64
5.6	Discussion	67
6	Compress and Forward with Optimized Quantization	69
6.1	Information Bottleneck Method	70

CONTENTS	XI
6.1.1 Information Rate Function	70
6.1.2 Relay Channel Adapted Information Rate Function	74
6.2 Modified Simulation Setup for Compress and Forward	79
6.2.1 Processing of Compression Indices in Soft Decoders	79
6.2.2 Rate Allocation	80
6.2.3 Complexity	81
6.3 Achievable Rates for Compress and Forward with Optimized Quantization . . .	82
6.3.1 Comparison to Decode and Forward	82
6.3.2 Comparison of Non-Orthogonal and Orthogonal Scheme	85
6.4 Discussion	87
7 Quantizer Design for Compress and Forward over Multiple Carriers	89
7.1 Preliminaries	90
7.1.1 Factor Graphs	92
7.1.2 Information Bottleneck Graphs	94
7.2 Quantizer Design for Perfect Channel Knowledge	96
7.2.1 Implicit Equalization with Partial Side Information	96
7.2.2 Explicit Equalization with Partial Side Information	98
7.2.3 Explicit Equalization with Full Side Information	99
7.2.4 Complexity	100
7.3 Quantizer Design for Imperfect Channel Knowledge	101
7.3.1 Implicit Channel Estimation and Equalization	101
7.3.2 Explicit Channel Estimation and Equalization	103
7.3.3 Complexity	104
7.4 Achievable Rates for Compress and Forward Using Multiple Carriers	104
7.4.1 Perfect Channel Knowledge	105
7.4.2 Imperfect Channel Knowledge	107
7.5 Discussion	109
8 Conclusion	111
Bibliography	115
Appendix	121
A Entropy and Mutual Information	121
A.1 Entropy	121
A.2 Joint and Conditional Entropy	122
A.3 Mutual Information for Discrete Random Variables	122
A.4 Differential Entropy	123

A.5	Relative Entropy	124
A.6	Mutual Information for Continuous Variables	124
A.7	Differential Entropy of Normal Distribution	125
B	Random Binning as in Wyner-Ziv Coding	125
C	Achievable Rates for Discrete Input	127
C.1	AWGN Channel	127
C.2	Max Flow Min Cut Sets for Relay Channel	127
D	Log-Likelihood-Ratios in a Turbo Decoder	128
E	Superposition Multilevel Coding - ρ for Parallel Detection	129
E.1	Complex Group Size $G = 6$:	130
E.2	Complex Group Size $G = 4$:	137
F	Information Bottleneck Algorithms	141
G	Joint Decoding for Compress and Forward	143
H	Explicit Equalization with Effective SNR and Full Side Information	144

Acronyms

A/D	Analog to Digital
AF	Amplify-and-Forward
ASK	Amplitude Shift Keying
AWGN	Additive White Gaussian Noise
BCJR	Bahl Cocke Jelinek Raviv
BER	Bit Error Rate
BICM	Bit Interleaved Coded Modulation
BP	Belief Propagation
BPSK	Binary Phase Shift Keying
cdf	cumulative distribution function
CF	Compress-and-Forward
DF	Decode-and-Forward
EXIT	Extrinsic-Information-Transfer
FD	Full-Duplex
FER	Frame Error Rate
FFT	Fast Fourier Transform
GI	Guard Interval
HD	Half-Duplex
I	Inphase
i.i.d.	independent and identical distributed
i.u.d.	independent and uniformly distributed

IB	Information Bottleneck
IBG	Information Bottleneck Graph
ICI	Inter-Carrier-Interference
IFFT	Inverse Fast Fourier Transform
IR	Incremental Redundancy
ISI	Inter-Symbol-Interference
LDPC	Low Density Parity Check
LLR	Log Likelihood Ratio
LTE	Long Term Evolution
LUT	Look-Up-Table
MAC	Multiple Access
MAP	Maximum-A-Posteriori
MARC	Multiple Access Relay Channel
MCS	Modulation and Coding Scheme
MFMC	Max-Flow Min-Cut
ML	Maximum Likelihood
MOE	Maximum Output Entropy
MRC	Maximum Ratio Combining
NSC	Non-Recursive Systematic Convolutional
OFDM	Orthogonal Frequency Division Multiplexing
P2P	Point-to-Point
PAPR	Peak to Average Power Ratio
pdf	probability density function
PLL	Phase Locked Loop
pmf	probability mass function
PSK	Phase Shift Keying
Q	Quadrature
QAM	Quadrature Amplitude Modulation
QPSK	Quarternary Phase Shift Keying
RSC	Recursive Systematic Convolutional

S/P	Serial-to-Parallel
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio
SPC	Single Parity Check
UMTS	Universal Mobile Telecommunications System

Symbols

S, R, D	source, relay, and destination node
u, u_1, u_2	information bits for p2p, 1st, and 2nd source message
$\hat{u}, \hat{u}_1, \hat{u}_2$	decoded bits for p2p, 1st, and 2nd source message
c, c_1, c_2	code bits for p2p, 1st, and 2nd source message
$X, x \in \mathcal{X}$	channel input (transmit signal)
$X_{S1}, x_{S1} \in \mathcal{X}_{S1}$	source transmit signal in 1st time slot
$X_{S2}, x_{S2} \in \mathcal{X}_{S2}$	source transmit signal in 2nd time slot
$\tilde{X}_{S2}, \tilde{x}_{S2} \in \tilde{\mathcal{X}}_{S2}$	source transmit signal in 2nd time slot with new information (only DF)
$X_R, x_R \in \mathcal{X}_R$	relay transmit signal (2nd time slot)
$X^p, x^p \in \mathcal{X}^p$	pilot signal
P, P_{S1}, P_{S2}, P_R	average transmit power of respective trasmit signals
$N, n \sim \mathcal{CN}(0, 1)$	Additive White Gaussian Noise (AWGN)
$N_R, n_R \sim \mathcal{CN}(0, 1)$	AWGN at relay (1st time slot)
$N_{D1}, n_{D1} \sim \mathcal{CN}(0, 1)$	AWGN at destination in 1st time slot
$N_{D2}, n_{D2} \sim \mathcal{CN}(0, 1)$	AWGN at destination in 2nd time slot
$d, d_{SR}, d_{SD}, d_{RD}$	distance between respective nodes
α	path-loss exponent
$H, \mathbf{h}, h \in \mathcal{H}$	general channel coefficient
$\hat{H}, \hat{h} \in \hat{\mathcal{H}}$	estimated channel coefficient
$H_{SR}, h_{SR} \in \mathcal{H}_{SR}$	channel coefficient on $S \rightarrow R$ link
$H_{SD}, h_{SD} \in \mathcal{H}_{SD}$	channel coefficient on $S \rightarrow D$ link
$H_{RD}, h_{RD} \in \mathcal{H}_{RD}$	channel coefficient on $R \rightarrow D$ link
$\hat{h}_{SR}, \hat{h}_{SR} \in \hat{\mathcal{H}}_{SR}$	estimated channel coefficient on $S \rightarrow R$ link
$Y, y \in \mathcal{Y}$	channel output (receive signal)

$Y_R, y_R \in \mathcal{Y}_R$	receive signal at relay (1st time slot)
$Y_{D1}, y_{D1} \in \mathcal{Y}_{D1}$	receive signal at destination in 1st time slot
$Y_{D2}, y_{D2} \in \mathcal{Y}_{D2}$	receive signal at destination in 2nd time slot
$Y^p, y^p \in \mathcal{Y}^p$	received pilot signal
$p(\cdot)$	probability density function
$\Pr\{\cdot\}$	probability mass function
σ^2	variance of corresponding random variable
μ	mean of corresponding random variable
$E\{\cdot\}$	expected value
$ \cdot $	magnitude or cardinality iff applied to set
$I(\cdot)$	mutual information
$H(\cdot)$	entropy
$h(\cdot)$	differential entropy
C	channel capacity
m	modulation order in bit/symbol (number of bit levels)
M	number of points in a signal space according to a modulation scheme
$\alpha_n, \alpha_l, \beta_n, \gamma_n$	amplitude of bit level n or l with respect to power P, P_{S1}, P_{S2}, P_R
α'_n	amplitude of virtual bit level n in superposition multilevel coding for DF
L	number of groups regarding power allocation of superposition modulation
G	group size regarding power allocation of superposition modulation
R_c	code rate
R_b	transmission rate of modulation and coding scheme
\mathbf{P}	puncturing matrix
L_C	memory of a convolutional code
$L(\cdot)$	log-likelihood ratio
$\Phi, \phi \in [0, 2\pi]$	phase error according to Phase Locked Loop (PLL)
γ	SNR of PLL (loop SNR)

List of Figures

2.1	Point-to-point transmission channel.	6
2.2	Capacity of Additive White Gaussian Noise (AWGN) channel for complex valued input.	8
2.3	Signal spaces for M -QAM with $M \in \{4, 16\}$	8
2.4	Capacity of AWGN channel with QAM input.	9
2.5	Twofold Parallel Concatenated Turbo Encoder.	10
2.6	Convolutional encoder with feedforward generator $(15)_8$ and feedback generator $(13)_8$	11
2.7	Iterative Turbo Decoding.	12
2.8	Bit level capacities of 16-QAM.	13
2.9	Multilevel encoder.	13
2.10	Multistage decoder.	13
2.11	Binary Ungerböck Partitioning.	14
2.12	Superposition modulation with multiple codes.	15
2.13	Superposition modulation with single code.	15
2.14	Signal space distribution of superposition modulation for equal power allocation with $m = 12$	16
2.15	Signal space distribution of superposition modulation for equal power allocation with $m = 18$	17
2.16	Signal space distribution of superposition modulation for unequal power allocation with $m = 3$	17
2.17	Signal space distribution of superposition modulation for grouped power allocation with $m = G \cdot L = 2 \cdot 4 = 8$ and $H(X) = 4.7159$	18
2.18	Signal space distribution of superposition modulation for grouped power allocation with $m = G \cdot L = 3 \cdot 4 = 12$ $H(X) = 5.0345$	19
2.19	Signal space distribution of superposition modulation for grouped power allocation with $m = G \cdot L = 6 \cdot 4 = 18$ and $H(X) = 5.0724$	19
2.20	Capacity of AWGN channel for superposition modulation with different power allocation schemes.	20

2.21	Blockdiagram of Bit Interleaved Coded Modulation.	21
2.22	Achievable rates R_b versus Signal to Noise Ratio (SNR) for BER = 10^{-5} for AWGN channel.	22
2.23	Sinc spectrum of each subcarrier $G_{Tx}(\cdot)$ causes no ICI with appropriate carrier spacing Δf around carrier frequency f_0	23
2.24	Blockdiagram of Orthogonal Frequency Division Multiplexing.	24
3.1	3-node relay Channel.	28
3.2	Cuts for 3-node relay channel.	28
3.3	Achievable rates, upper capacity bound and optimal parameters ρ and τ versus source relay distance d for individual power constraint with path loss exponent $\alpha = 4$ and transmit power $P = 1$	39
3.4	Achievable rates, upper capacity bound and optimal parameters ρ and τ versus source relay distance d for uniform power allocation with pathloss exponent $\alpha = 4$ and sum energy $E_{sum}^{non} = 1$	41
3.5	Achievable rates, upper capacity bound and optimal parameters ρ and τ versus source relay distance d with pathloss exponent $\alpha = 4$ and sum energy $E_{sum}^{non} = 1$	43
3.6	Achievable rates, upper capacity bound and optimal parameters ρ and τ versus source relay distance d with pathloss exponent $\alpha = 4$, sum energy $E_{sum}^{non} = 1$, and $\beta = 0.5$	44
4.1	Conventional processing at source and relay for the Multiple Access (MAC) phase. Due to error free decoding, the relay has complete knowledge about u_1	48
4.2	Concept of Distributed Superposition Multilevel Coding for the classical relay channel.	48
4.3	Achievable rates and optimal correlation coefficient ρ versus source relay distance d for DF for superposition modulation and QAM with $\alpha = 4$ and transmit power $P = 1$	53
5.1	Exact (cf. (5.3)) and approximated (cf. (5.6)) $p_\Phi(\phi)$ for different loop SNRs γ	57
5.2	$p_{Y_{D2}^{non} X_R\tilde{X}_{S2}}(r, \varphi)$ with $h_{SD} \sqrt{P_{S2}}\rho \cdot x_R = 5 \cdot e^{j\frac{\pi}{4}}$, $r_0 = 0$, $\varphi_0 = 0$ (cf. (5.9)) for exact (cf. (5.3)) and approximated (cf. (5.6)) $p_\Phi(\phi)$ and different loop SNRs γ	59
5.3	Achievable rates and optimal correlation coefficient ρ versus source relay distance d for DF with QAM input in the presence of phase errors with pathloss exponent $\alpha = 4$ and transmit power $P = 1$	63
5.4	Achievable DF rates versus SNR (sum energy) at $d = 0.2$ with $\alpha = 4$ including imperfect phase knowledge dependent on γ . Distributed beamforming with $\rho = 1$ is optimal in the whole range.	65

5.5	Achievable DF rates versus SNR (sum energy) at $d = 0.4$ with $\alpha = 4$ including imperfect phase knowledge dependent on γ	66
6.1	Relay Processing for CF.	69
6.2	Exemplary information-rate curve for $x \in \mathcal{X} = \{-1, 1\}$, $y = x + n$, $n \in \mathcal{N}(0, 1)$, $ \mathcal{Z} = 64$ and $\epsilon = 10^{-4}$	71
6.3	Information-rate curve for different cardinalities $ \mathcal{Z}_1 = 2$, $ \mathcal{Z}_2 = 3$, $ \mathcal{Z}_3 = 4$, and $ \mathcal{Z}_4 = 16$ with $x \in \mathcal{X} = \{-1, 1\}$, $y = x + n$, $n \in \mathcal{N}(0, 1)$, and $\epsilon = 10^{-4}$	73
6.4	3-node relay channel with CF.	74
6.5	Different quantizer rates r versus SNR (direct link of the 3-node relay channel) comparing deterministic quantization ($r = H(\hat{Y}_R Y_{D1})\forall \hat{\mathcal{Y}}_R \in \{2, \dots, 16\}$) to random quantization ($r \in \{\frac{r_{max}}{6}, \dots, r_{max}\}$) where r_{max} corresponds to $ \hat{\mathcal{Y}}_R = 16$ (red and blue match).	78
6.6	Block diagram for CF transmission with joint destination processing.	79
6.7	Achievable rates R_b^{xs1} versus SNR on direct $S \rightarrow D$ link for $ \hat{\mathcal{Y}}_R = 8$ and $d = 0.8$ compared to achievable rates for a direct transmission. Different colors indicate different modulation orders.	81
6.8	Total achievable rates for CF and DF versus sum SNR (sum energy) at $d = 0.8$ for non-orthogonal channel access in the MAC phase.	82
6.9	Total achievable rates for CF and DF versus SNR (sum energy) at $d = 0.9$ for non-orthogonal channel access in the MAC phase.	83
6.10	Total achievable rates for CF and DF versus SNR (sum energy) at $d = 0.8$ for orthogonal channel access in the MAC phase.	84
6.11	Total achievable rates for CF and DF versus SNR (sum energy) at $d = 0.9$ for orthogonal channel access in the MAC phase.	84
6.12	Total achievable rates for CF versus SNR (sum energy) at $d = 0.8$ comparing orthogonal to non-orthogonal channel access in the MAC phase.	85
6.13	Total achievable rates for CF versus SNR (individual power) at $d = 0.8$ comparing orthogonal to non-orthogonal channel access in the MAC phase.	86
7.1	Relay Processing for CF with direct quantization of y_R exploiting observed pilot y_R^p	91
7.2	Exemplary factor Graph.	92
7.3	Factor graph of Fig. 7.2 with Marginalization.	93
7.4	Factor graph representation of the quantizer design for one subcarrier including the compression mapping of the quantizer itself.	93
7.5	Transformation of factor graph into simple IBG and further modification to consider side information in a Wyner-Ziv coding manner.	94
7.6	IBG representation of factor graphs in Fig. 7.4.	95

7.7	IBG with cascading into 3 stages considering part of side information.	97
7.8	IBG with explicit equalization considering part of the side information.	99
7.9	IBG with explicit equalization and cascading considering full side information.	100
7.10	IBG with cascading into 3 stages with implicit channel estimation and equalization.	102
7.11	IBG with explicit channel estimation and equalization.	103
7.12	Total achievable rates for CF versus SNR (individual power) on direct link at $d = 0.8$ with $\alpha = 4$ comparing different quantizer approaches for orthogonal channel access.	105
7.13	Total achievable rates for CF versus SNR (sum energy) on direct link at $d = 0.8$ with $\alpha = 4$ comparing non-orthogonal and orthogonal channel access for exploitation of total side information.	106
7.14	Total achievable rates for CF versus SNR (individual power) on direct link at $d = 0.8$ with $\alpha = 4$ comparing different quantizer approaches for orthogonal channel access.	108
A.1	Visualization of mutual information is difference of entropies.	123
A.2	probability density functions (pdfs) of y_R and y_{D1} for Binary Phase Shift Keying (BPSK) input with SNRs (3.9, 0dB) according to a line setup: Source, relay, and destination are at positions $d_S = 0$, $d_R = 0.8$, and $d_D = 1$ with pathloss exponent $\alpha = 4$, transmit power $P_{S1} = 1$ and noise power $N_0 = 1$	126
H.3	IBG for explicit equalization using different quantizers on each sub carrier depending on effective SNR.	144

List of Tables

2.1	Puncturing Patterns (octal)	11
4.1	Values of ρ for different combinations of K and L for $G = 6$ (respectively 3 layers in Inphase (I) and Quadrature (Q)), $d = 0.4$, and $P_{S1} = P_{S2} = P_R = P = 1$	52
7.1	Complexity of Quantizer Design for Perfect Channel Knowledge	100
7.2	Complexity of Quantizer Design for Imperfect Channel Knowledge	104
B.1	Simple Wyner-Ziv coding scheme using random binning.	127
E.2	Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.3$	130
E.3	Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.4$	131
E.4	Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.5$	132
E.5	Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.6$	133
E.6	Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.7$	134
E.7	Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.8$	135
E.8	Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.9$	136
E.9	Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.3$	137
E.10	Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.4$	137
E.11	Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.5$	138
E.12	Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.6$	138
E.13	Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.7$	139
E.14	Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.8$	139
E.15	Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.9$	140

Chapter 1

Introduction

1.1 Motivation and Objective

The 3-node relay channel is a well investigated subject in the literature: It was originally introduced as classical relay channel by van der Meulen in [vdM71, vdM77]. Since that time a lot of work has been done. In [CE79, KGG05] an upper bound on the capacity and the achievable rates of the relaying strategies Decode-and-Forward (DF) and Compress-and-Forward (CF) are derived for full-duplex relays by means of the Max-Flow Min-Cut (MFMC) theorem [FF56, FF62, CT91, EK12]. In [HMZ05, HM02], the authors consider a more practical scheme based on a half-duplex relay (cannot receive and transmit at the same time) so that the length of the arising two time slots has to be optimized [WH10a]. For DF, this optimization usually implies the use of different codes at source and relay in the two time slots which is known as DF with incremental redundancy as it generates a concatenated code for the destination decoder [ACLY00]. In contrast to this information theoretically natural concept, Laneman introduced a simplified scheme assuming repetition coding in [Lan02, LTW04] which requires time slots of equal length to apply maximum ratio combining. Throughout this work, the latter suboptimal scheme is not considered and, thus, DF is always meant to add incremental redundancy at the relay. A very detailed summary of these investigations can be found in [Wei12] which can be seen as preceding work of this thesis. Information theoretic analysis reveals that DF performs very close to the upper bound on capacity when the relay is placed considerably closer to the source than to the destination whereby, compared to a direct transmission, a huge gain can be achieved with the help of distributed beamforming [MBM07].¹ As distributed beamforming requires that the source adjusts the phase of its transmit signal such that it superposes constructively with the relay's signal at the destination, phase errors will degrade the performance. So far such phase errors have not been considered but only the extreme cases

¹For the degraded relay channel, DF is known to achieve channel capacity [CE79, KGG05].

of perfect phase knowledge or uniformly distributed phase fading [KGG05]. This work will analyze the influence of such phase errors, which cannot be avoided in practice.

More precisely, a nonuniformly distributed random variable will be introduced. Furthermore, the influence of practical coding providing only discrete rates due to limited Modulation and Coding Schemes (MCSs) as well as concepts like distributed multilevel coding will be investigated in combination with superposition modulation [KK15b, KK15c]. Similar as in [Wei12], general non-orthogonal and pragmatic orthogonal Multiple Access (MAC) will be distinguished in a two time slot scheme which is required to realize half-duplex relaying.

For scenarios, where the relay is very close to the destination and far from the source, DF is outperformed by CF because the source-relay link becomes the bottleneck of the system. Unfortunately, also CF leaves a gap to the upper bound on the capacity except for relay and destination being at the same position. Still, there is no better relaying strategy known until now according to the literature. Therefore, CF is investigated from a more practical point of view. Most importantly, the question of how to implement the compression at the relay is addressed: One suitable possibility is to apply quantization whereby an optimal quantizer can be designed with the help of the Information Bottleneck (IB) method [TPB99, Slo02, Zei12]. Again, specific MCSs are used to investigate the influence of practical coding with only discrete rates [KK17a, KK17b]. Finally, the quantizer design is extended for transmission over multiple carriers to tackle frequency selective fading channels [KK17c]. As this cannot be solved feasibly using the IB method, suboptimal quantizer design approaches are derived using Information Bottleneck Graphs (IBGs) [LSB16].

1.2 Thesis Outline

The thesis is outlined as follows: Ch. 2 sketches basic concepts of communication theory to follow the analysis of the main chapters focusing on the 3-node relay channel which is introduced and investigated from an information theoretic perspective in Ch. 3. The following chapters, Ch. 4 and Ch. 5, investigate DF from a more practical perspective considering multilevel coding with superposition modulation and Bit Interleaved Coded Modulation (BICM) with imperfect synchronization. Finally, Ch. 6 and Ch. 7 are concerned with quantizer design for CF considering different channel models (single carrier with path-loss and multi carrier with Rayleigh fading) before Ch. 8 concludes the thesis.

1.3 Notation

Throughout this thesis, lower case letters are used to denote realizations, e.g. $x \in \mathcal{X}$, of random variables denoted with capital letters, e.g. X , whereby calligraphic letters, e.g. \mathcal{X} denote generally sets or alphabets with cardinality $|\mathcal{X}|$. The probability density function (pdf) of a continuous random variable X is $p_X(x)$. Whenever the relation of argument and random variable is clear from the context, notation will be simplified to $p(x)$ for the sake of clarity. For discrete random variables, the corresponding distribution is given by a probability mass function (pmf) which is denoted by $\Pr\{X = x\} = \Pr\{x\}$. Joint and conditional pdfs or pmfs will be denoted by $p(x, y)$, $p(y|x)$, $\Pr\{x, y\}$, and $\Pr\{y|x\}$. Specific distribution functions like Gaussian, complex Gaussian, and Rayleigh are denoted by $\mathcal{N}(\mu, \sigma^2)$, $\mathcal{CN}(\mu, \sigma^2)$, and $\mathcal{R}(\sigma^2)$ whereby μ and σ^2 denote respective mean and variance. For complex Gaussians, the variance is divided equally on real and imaginary part. Transmit and receive signals are usually denoted by x and y with respective subscripts for indicating affected node and time slot whereby transmit signals usually have unit variance $\sigma_X^2 = \mathbb{E}\{|X|^2\} = 1$. If not stated otherwise, notation refers to complex baseband whereby real and imaginary part of a variable, e.g., x are indicated by primes, e.g., x' and x'' . Usually, variables are scalar while vectors and matrices are highlighted using lower case bold letters and capital bold letters, respectively. The identity matrix for example is denoted by \mathbf{I} .

Chapter 2

Fundamentals - Point to Point Transmission

As Point-to-Point (P2P) transmissions are part of a relay network, this chapter is concerned with basic components and principal variables necessary for digital communication. Firstly, the channel model and its information theoretic capacity will be introduced in Sec. 2.1 and Sec. 2.2. Afterwards, Sec. 2.3 describes state-of-the-art Modulation and Coding Schemes (MCSs) including Quadrature Amplitude Modulation (QAM), multilevel coding, superposition modulation, turbo coding and Bit Interleaved Coded Modulation (BICM). Sec. 2.4 presents the simulation setup and results for a range of practically relevant (discrete) rates according to different MCSs. Finally, as multi-carrier transmission is well known to combat frequency selective fading channels, Orthogonal Frequency Division Multiplexing (OFDM) will be briefly introduced in Sec. 2.5. Please note that OFDM is only relevant in Ch. 7 for Compress-and-Forward (CF) where both perfect and imperfect channel estimation (cf. Sec. 2.5.2) are considered.

2.1 Channel Model

This section describes the channel model used throughout this thesis. At first, the Additive White Gaussian Noise (AWGN) channel model including a path loss is used to keep the analysis simple. Usually, gained insights can be straightforwardly generalized to a more realistic channel model considering frequency selective fading. The state-of-the-art technique to combat frequency selectivity is to use multi-carrier systems like OFDM. As it will be illustrated in Sec. 2.5, a frequency selective wide-band channel is decomposed into narrow-band flat Rayleigh fading channels. As the quantizer design cannot be easily applied to multi-carrier transmission, Ch. 7 investigates such quantizer design in detail using the Rayleigh fading channel model described in Sec. 2.1.2.

2.1.1 AWGN Channel with Path Loss

The AWGN channel is a very common and simple model for P2P transmissions whereby input $x \in \mathcal{X}$ and output $y \in \mathcal{Y}$ are drawn from random variables X and Y . As its name implies, the receive signal y is obtained by adding noise $n \sim \mathcal{CN}(0, \sigma_N^2)$ to transmit signal x . The term white refers to a constant power spectral density. As this work considers a relay network where the



Figure 2.1: Point-to-point transmission channel.

relative link quality plays an important role, the simple AWGN channel model is extended by an attenuating constant channel coefficient h as illustrated in Fig. 2.1.

$$y = \sqrt{P}hx + n \quad (2.1)$$

If not stated otherwise, the channel coefficient $h = d^{-\frac{\alpha}{2}}$ only represents the path loss depending on the distance d and the path loss exponent α which is usually in the range $[2, 5]$ (free space attenuation up to dense urban environments). The variance $\sigma_x^2 = 1$ of transmit signal x is normalized to unity because the transmit power is denoted by P . Furthermore, the noise power $\sigma_N^2 = 1$ is also normalized to unity so that the Signal to Noise Ratio (SNR) at the receiver is

$$\text{SNR} = |h|^2 P. \quad (2.2)$$

2.1.2 Rayleigh Fading

The above mentioned channel model is usually appropriate for free space propagation without any obstacles but not suitable for the most terrestrial environments. A more realistic channel model is the Rayleigh fading channel which depends on a random channel coefficient h instead of a deterministic one as before. A statistical channel coefficient instead of a deterministic one is meaningful due to its large number of unpredictable parameters like amplitude, phase, and angle of arrival arising from multiple reflective propagation paths. For example, a signal transmitted from a base station may be impaired by multiple reflections and diffractions at buildings or other obstacles nearby before reaching the receiver. Fortunately, due to a usually large number of propagation paths, the central limit theorem can be applied. Hence, a complex channel coefficient $h \in \mathcal{H}$ can be drawn from a Gaussian distribution, $h \sim \mathcal{CN}(0, \sigma_H^2 = d^{-\alpha})$ where $\sigma_H^2 = d^{-\alpha}$ is the path loss dependent average channel gain. As its name implies, the magnitude $|h|$ of the channel coefficient follows a Rayleigh distribution $\mathcal{R}(d^{-\alpha})$. The channel model (2.1) is only used with this random coefficient to model subchannels of the OFDM system in Ch. 7. As

such subchannels are correlated over frequency, a vector of channel coefficients (denoting the channel impulse response) is randomly generated in time domain and then Fourier transformed such that the above statistics are fulfilled in frequency domain. The channel impulse response is modeled as a tapped delay line with statistical independent taps and a uniform power delay profile. The length of the channel impulse response, i.e., the number of taps, coincides with the length of the guard interval (cf. Sec. 2.5 and Ch. 7).

2.2 Channel Capacity

According to Shannon's information theory [Sha48, CT91], the channel capacity C is defined as the maximum mutual information $I(X; Y)$ between input and output with respect to the input distribution $p(x)$ under an average power limitation.

$$C = \max_{p(x): E\{|X|^2\} \leq 1} I(X; Y) \quad (2.3a)$$

A definition of mutual information by means of entropy as a statistical measure is given in App. A. For the Gaussian channel as depicted in Fig. 2.1, the capacity can be derived in closed form as shown in the following [CT91].

$$\begin{aligned} I(X; Y) &= h(Y) - h(Y|X) \\ C &= \log_2 \left(\pi e (|h|^2 P + 1) \right) - \log_2 \left(\pi e \sigma_N^2 \right) \\ &= \log_2 \left(1 + |h|^2 P \right) \end{aligned} \quad (2.3b)$$

From (2.3b) it becomes clear that the capacity

$$C(\text{SNR}) = \log_2(1 + \text{SNR}) \quad (2.3c)$$

is a function of $\text{SNR} = |h|^2 P$ as plotted in Fig. 2.2 and serves as an upper bound for all (discrete) modulation schemes.

2.3 Modulation and Coding Schemes

In addition to an information theoretic analysis with continuous rates and Gaussian alphabets, practical Modulation and Coding Schemes (MCSs) supporting only discrete rates will be considered. Therefore, state-of-the-art coded modulation, involving Quadrature Amplitude Modulation (QAM) and turbo coding, is used, whereby the question arises how coding and modulation shall be combined in order to approach the channel capacity for Gaussian input. On one

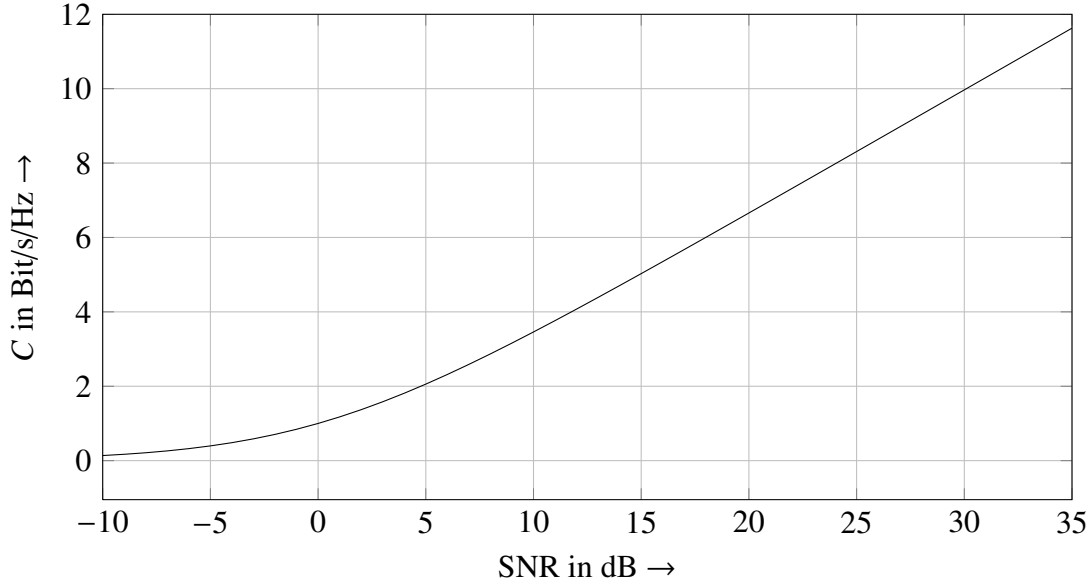
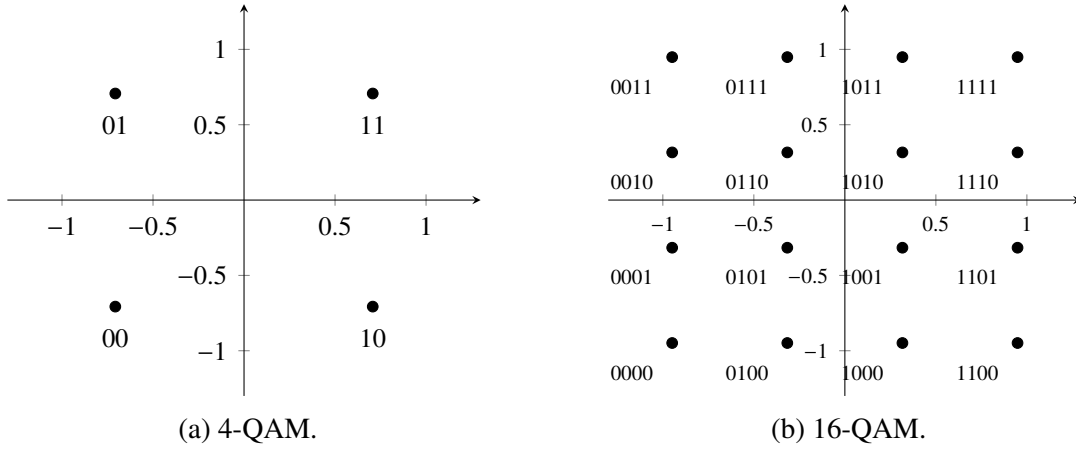


Figure 2.2: Capacity of AWGN channel for complex valued input.

Figure 2.3: Signal spaces for M -QAM with $M \in \{4, 16\}$.

hand, multilevel coding and superposition modulation are very suitable but costly to implement. On the other hand, Bit Interleaved Coded Modulation (BICM) is better suited for practical implementation due to its simple structure while performance is only slightly degraded.

2.3.1 Quadrature Amplitude Modulation

The practically most relevant modulation scheme for wireless communications, especially for very high data rates, is QAM because it very efficiently uses the complex plane for limited average and maximum transmit power. The signal spaces of QAM are shown in Fig. 2.3 for two selected modulation orders.¹ Usually, an M -QAM uses M signal points that correspond

¹Please note that $\sigma_X^2 = 1$ since the transmit power is denoted by P

to $m = \log_2 M$ bit/symbol. As its name implies, it can be made up quadratically from two Amplitude Shift Keyings (ASKs) in real and imaginary part also known as Inphase (I) and Quadrature (Q) component. Hence, it can be easily mapped and demapped independently in I and Q, i.e., half of the bit/symbol for each component. Throughout this work, gray mapping is applied to ensure only one bit error for mixing up neighboring symbols. Therefore, Fig. 2.3 shows in addition an exemplary mapping for 4- and 16-QAM. The achievable rates and the loss to Shannon's capacity are depicted in Fig. 2.4. The numerical evaluation of the mutual information terms for different modulation orders is sketched in App. C.1. From Fig. 2.4 it is obvious that a discrete modulation scheme with finite cardinality cannot achieve the capacity (cf. (2.3b)). In the low SNR regime however, the gap to capacity is very small. In the high SNR regime, the final rate (mutual information) is saturated into the entropy of the input alphabet.

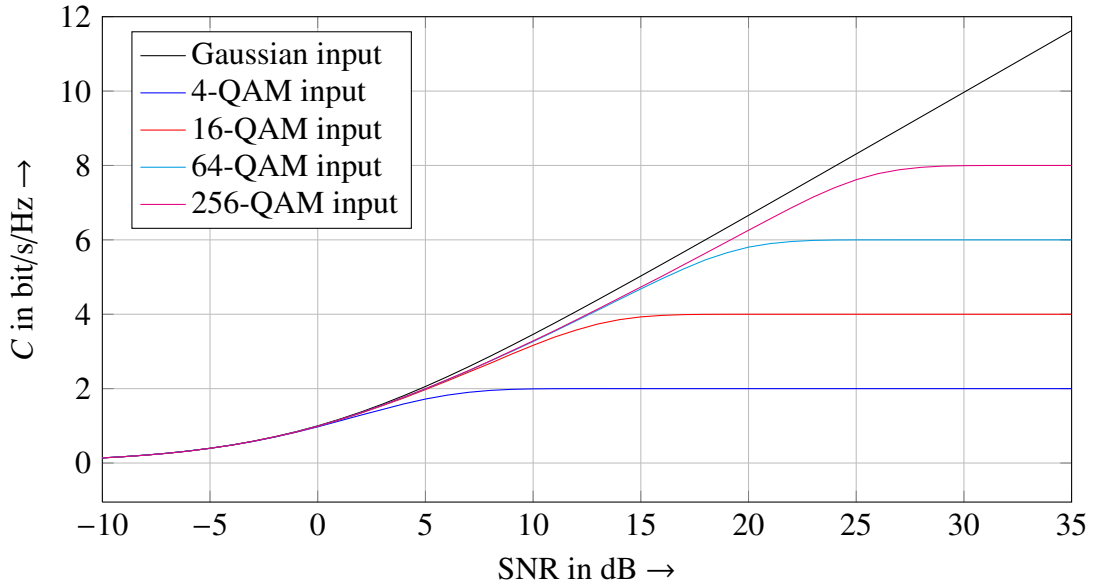


Figure 2.4: Capacity of AWGN channel with QAM input.

2.3.2 Turbo Coding

Due to transmission errors introduced by the channel, channel coding is necessary to approach channel capacity. Turbo codes are very powerful state-of-the-art codes besides others like Low Density Parity Check (LDPC) codes [Gal63, Sho03] that are implemented in current mobile communication systems like the Universal Mobile Telecommunications System (UMTS) and the Long Term Evolution (LTE). The key idea is to create very powerful codes by concatenation of simple codes [BGT93, NFK07]. As a consequence of this concatenation, optimal maximum likelihood decoding is usually not feasible and, thus, replaced by an iterative turbo decoding process exchanging extrinsic information between component decoders. Strictly speaking turbo codes are known to be parallel concatenated convolutional codes as invented by Berrou et al. in

1993 [BGT93]. However, the term is also used to emphasize the turbo fashion of the decoding principle of such codes, i.e., all kinds of concatenations (component codes, mappers, etc.) exchanging extrinsic information form a turbo code. In a wide sense even LDPC codes can be understood as turbo codes due to their complex concatenation of repetition and Single Parity Check (SPC) codes. The code construction, that is, the choice of component codes and number of iterations, can be done with the help of Extrinsic-Information-Transfer (EXIT) charts since the initial work of Stefan ten Brink [tB01a, tB01b].

Encoding Structure

The popularity of turbo codes is principally reasoned in their simple structure as depicted in Fig. 2.5 which shows two parallel concatenated encoders. Theoretically, this structure could be extended to arbitrarily many concatenated codes which is however not relevant for practical implementation. Due to the interleaver in Fig. 2.5 the component encoders C_1 and C_2 will pro-

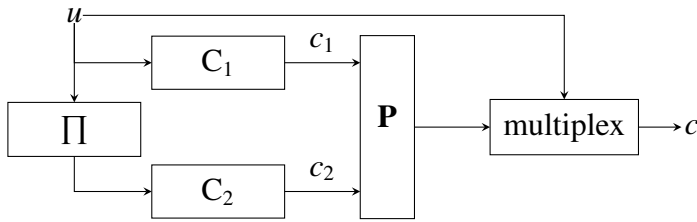


Figure 2.5: Twofold Parallel Concatenated Turbo Encoder.

duce different parity check bits even when the same encoding structure is used for both. The component codes are usually Recursive Systematic Convolutional (RSC) codes with a moderate memory of $L_c \in \{3, 4, 5\}$. Due to the inefficiency of transmitting the systematic part multiple times, encoders deliver only parity bits while information bits have a separate path. The puncturing matrix \mathbf{P} can finally puncture bits to adapt the code rate

$$R_c = \frac{R_c^1 \cdot R_c^2}{R_c^1 + R_c^2 - R_c^1 \cdot R_c^2} \quad (2.4)$$

where R_c^1 and R_c^2 are the code rates of the component codes.

In this work, the component encoders are chosen as depicted in Fig. 2.6 which leads to the well known turbo code with $R_c = \frac{1}{3}$ specified in UMTS and LTE [NFK07]. Feedforward and feedback generator of the convolutional code are given in octal notation: $(15)_8$ and $(13)_8$. To be adaptive to varying channel conditions, 8 different code rates are obtained via puncturing as described in Table 2.1 [LCG⁺09].

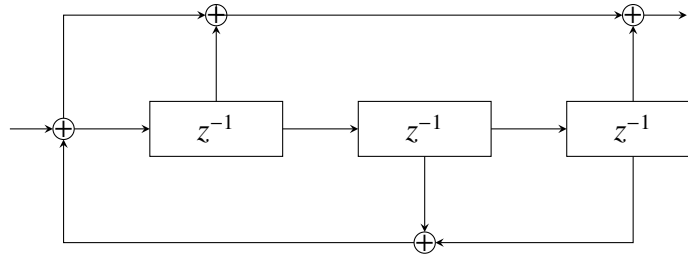


Figure 2.6: Convolutional encoder with feedforward generator $(15)_8$ and feedback generator $(13)_8$.

Table 2.1: Puncturing Patterns (octal)

4/5	2/3	4/7	1/2	4/9	2/5	4/11	1/3
100	101	101	121	125	125	335	377
001	021	261	263	363	377	377	377

Turbo Decoding

Turbo decoders exchange information about the bits u_l of an information word \mathbf{u} , usually in the form of Log Likelihood Ratios (LLRs). The detailed derivation for the LLR of interest in a turbo decoder is given in App. D. Assuming a systematic encoding, the final result is

$$L(u_l) = L(y_l|u_l) + L_a(u_l) + \log \frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \prod_{i=0, i \neq l}^{n-1} e^{-(L(y_i|c_i))c_i} \prod_{j=0, j \neq l}^{k-1} e^{-L(u_j)u_j}}{\underbrace{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \prod_{i=0, i \neq l}^{n-1} e^{-(L(y_i|c_i))c_i} \prod_{j=0, j \neq l}^{k-1} e^{-L(u_j)u_j}}_{L_e(u_l)}}. \quad (2.5)$$

Investigating (2.5), one can see that the LLR $L(u_l)$ of a specific information bit u_l is additively composed of three shares:

1. $L(y_l|u_l)$ is directly known due to channel observation (always intrinsic),
2. $L_a(u_l)$ is a priori knowledge which can also arise extrinsically from another decoder (turbo principle),
3. $L_e(u_l)$ is extrinsic knowledge which is gained from other code bits (generated by decoder D) and, thus, can be used as a priori knowledge in another decoder (turbo principle).

The resulting turbo decoding principle is visualized in Fig. 2.7 where the subtraction cancels the amount that is already known in the respective component decoder. Hence, only extrinsic information is exchanged. The component decoders D_1 and D_2 are usually Bahl Cocke Jelinek Raviv (BCJR) decoders which evaluate (2.5) efficiently based on trellis diagrams [BCJR74].

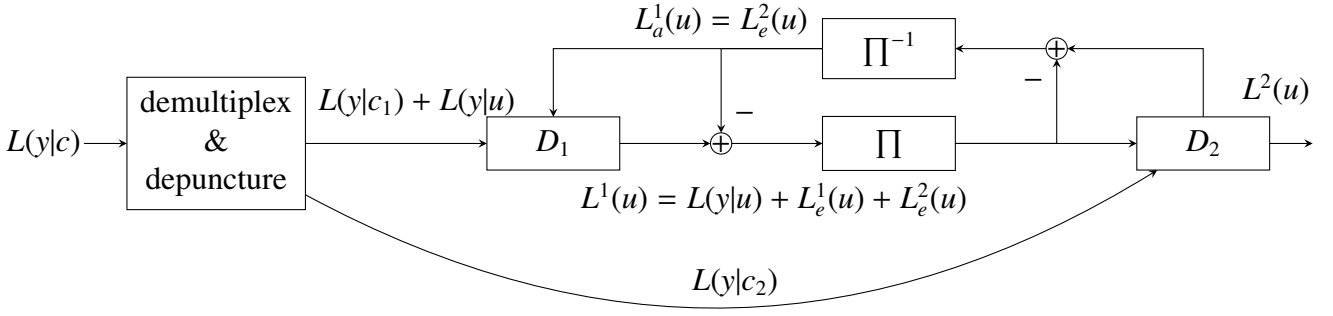


Figure 2.7: Iterative Turbo Decoding.

2.3.3 Multilevel Coding

The combination of higher order modulation and channel coding is known as coded modulation. Usually, different bit levels are not independent after transmission through the channel. One exception is 4-QAM, which has $m = 2$ bit/symbol, however, the two bits are mapped independently in I and Q. There are different possibilities to implement coded modulation, e.g., trellis coded modulation, multilevel coding, or BICM. In this work, the focus is set to multilevel coding and BICM for the following two reasons. Firstly, multilevel codes are very useful in combination with superposition modulation to exploit the advantages of Decode-and-Forward (DF) with non-orthogonal channel access (cf. Ch. 5). Secondly, BICM as introduced in Sec. 2.3.5, is generally preferable for practical implementation due to only small performance loss compared to optimal coded modulation while its implementation simplifies tremendously.

The general idea of multilevel coding, originally introduced by Imai in [IH77], is to encode each bit level independently according to its respective equivalent channel capacity (unequal error protection) and follows directly from the chain rule of mutual information. For the case of a bijective mapping, which is lossless from an information theoretic perspective, the mutual information $I(X; Y)$ between transmit symbol $x \in \mathcal{X}$ and receive symbol $y \in \mathcal{Y}$ is equal to the mutual information $I(Y; C_1, C_2, \dots, C_m)$, where c_1, \dots, c_m are the different (encoded) bit levels of x . Following the chain rule of mutual information [CT91] directly leads to the decomposition of $I(X; Y)$ into equivalent channels for each bit level.

$$I(Y; C_1, C_2, \dots, C_m) = I(Y; C_1) + I(Y; C_2|C_1) + \dots + I(Y; C_m|C_1, C_2, \dots, C_{m-1}) \quad (2.6)$$

Therefore, the transmission of x over a channel delivering y can be interpreted as a successive transmission of bits $c_n, \forall n \in \{1, \dots, m\}$ over m virtual channels whereby the n th channel requires knowledge of previously transmitted bits c_1, \dots, c_{n-1} . The equivalent bit level capacities

according to (2.6) are plotted in Fig. 2.8.² The arising signal processing chains at transmitter

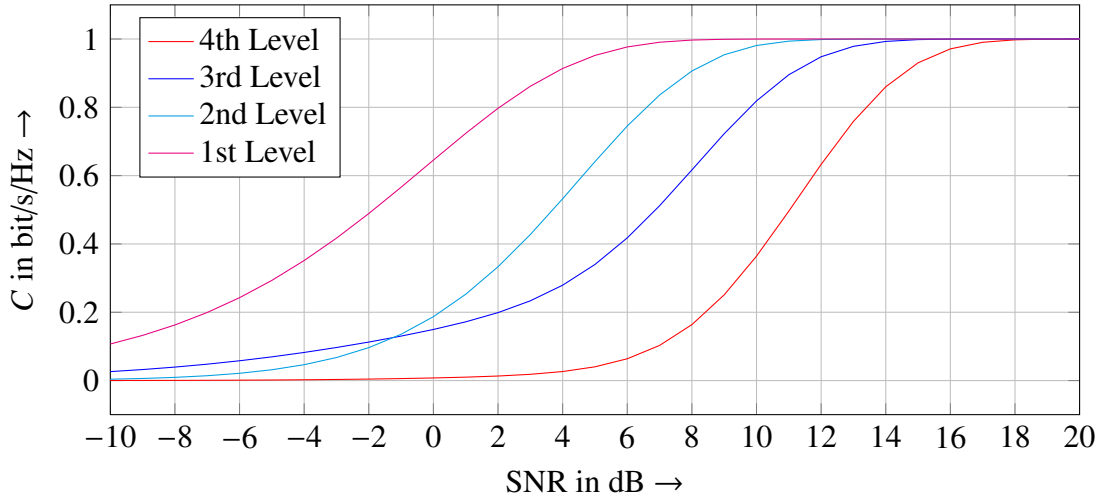


Figure 2.8: Bit level capacities of 16-QAM.

and receiver are depicted in Fig. 2.9 and Fig. 2.10. For the encoding, a sequence of information bits u is split into m different sequences of bits u_n (respective lengths depend on individual code rates) which are independently encoded into c_n . As induced by (2.6), decoding is implemented successively as depicted in Fig. 2.10 where the current stage exploits all previous decisions.

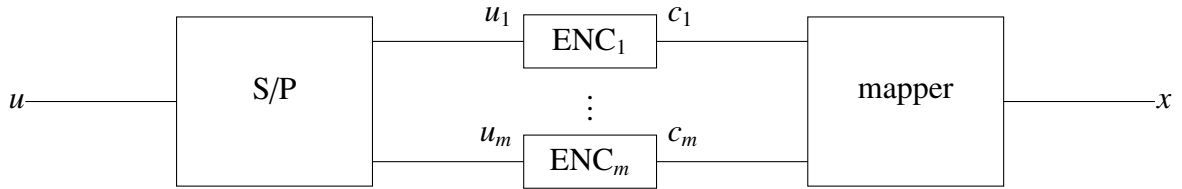


Figure 2.9: Multilevel encoder.

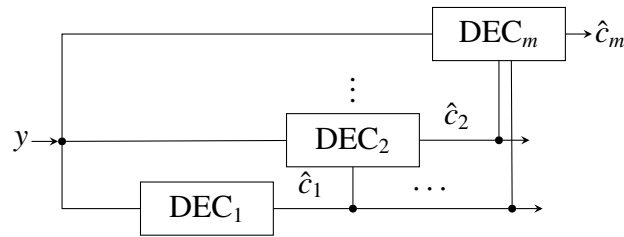


Figure 2.10: Multistage decoder.

The question arising in practice is how to obtain the different bit levels with individual code rates $R_c^n \leq I(Y; C_n | C_1, \dots, C_{n-1})$. Therefore, the set of all possible signal points (cf. Fig. 2.3) is separated step by step in dependence of the value of the bit in each level. There are different possibilities to do this set partitioning: natural partitioning, gray labeled partitioning, set

²The bit levels are partitioned according to Ungerböck as shown in Fig. 2.11.

partitioning according to Ungerböck as in Fig. 2.11 [Ung82], and more [WFH99]. Important to note is that all partitioning schemes are equally good in terms of sum capacity. The type of the set partitioning plays an important role in practical implementations. The partitioning of Ungerböck for example increases the minimum euclidean distance within each level as illustrated in Fig. 2.11 which leads to an increasing code rate along the levels so that high levels may be uncoded (cf. Fig. 2.8).

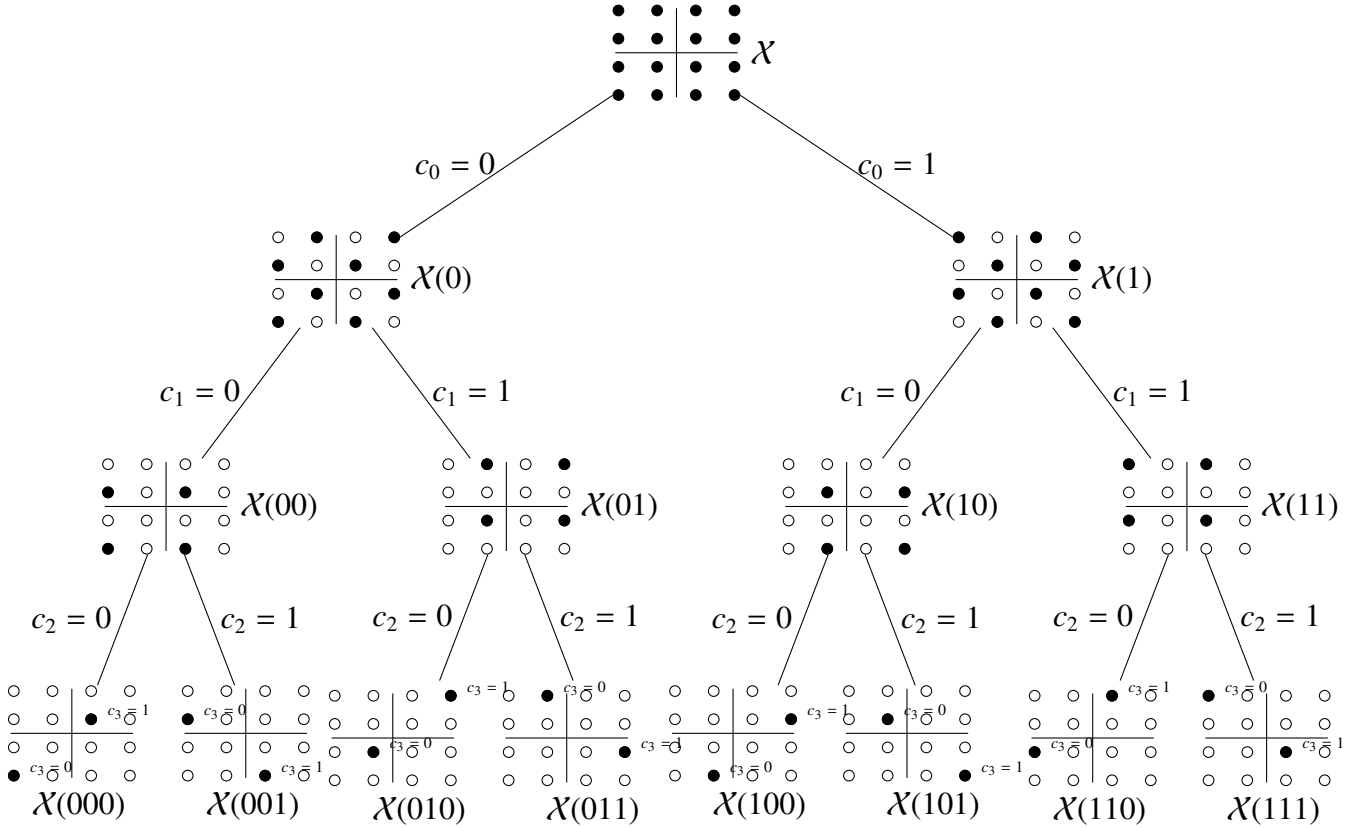


Figure 2.11: Binary Ungerböck Partitioning.

2.3.4 Superposition Modulation

A special case of multilevel coding is superposition modulation which achieves a natural shaping gain without active shaping [DRU97]. In contrast to classical multilevel coding where the modulation scheme is given and set partitioning applied, superposition modulation simply creates a weighted sum of m Binary Phase Shift Keying (BPSK) modulated bit levels c_n [WH10b]:

$$x = \sum_{n=1}^m \alpha_n e^{j\chi_n} \cdot c_n, \quad \text{with} \quad c_n \in \{-1, +1\} \quad (2.7)$$

where amplitude α_n and phase χ_n of each layer may be freely chosen with respect to a power constraint. This superposition of bit layers can create every possible complex symmetric signal constellation. Hence, schemes like ASK, Phase Shift Keying (PSK), or hybrids like QAM can be considered as special cases of superposition modulation [WH10b]. The motivation for the use of superposition modulation is usually the shaping gain, i.e., to close the gap to Shannon's capacity by creating a Gaussian like distribution. In fact, for an infinite number of levels, this distribution becomes exact and achieves the capacity. According to this goal, it is sufficient to consider only real valued superposition modulation ($\chi_n = 0$) because a complex Gaussian like distribution can be achieved by quadrature modulation [Wo11]. For the sake of completeness, there are irregular complex signal constellations as shown in [WH10b], however, their performance is always worse than using two real valued superposition modulations in I and Q [Wo11]. Figures 2.12 and 2.13 illustrate the encoding in combination with superposition modulation. The bit levels can be either encoded independently (cf. Fig. 2.12) as in classical

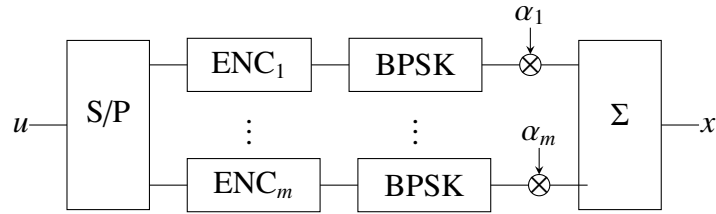


Figure 2.12: Superposition modulation with multiple codes.

multilevel coding or jointly using a single code (cf. Fig. 2.13). Adding an interleaver after the encoders will lead to a BICM scheme (cf. Sec. 2.3.5). At the cost of complexity, an individual encoding of the layers can improve the decoding convergence of a practical BICM scheme as shown in [TP10]. In relation to multilevel coding, the authors of [MP04, TP10] use the term Σ -mapping derived from the sum of individual BPSK mappings.

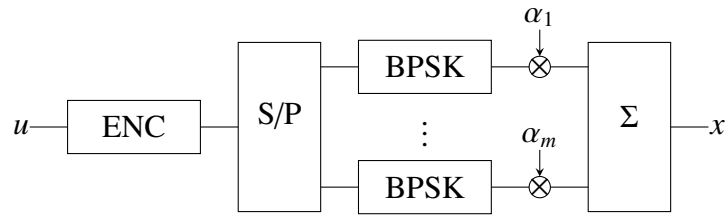


Figure 2.13: Superposition modulation with single code.

As mentioned before, the amplitude of each layer can be chosen arbitrarily, that is, α_n is chosen such that the power α_n^2 spent for each level n satisfies an average power constraint, e.g., $\sum_{n=1}^m \alpha_n^2 = 1$. Although any power allocation scheme is possible, there are some meaningful schemes to design signal spaces with desirable properties like equidistant symbols or Gaussian like distributions. According to [Wo11], equal, unequal, and grouped power allocation can be

distinguished whereby unequal power allocation is structured following the exponential law through the layers and grouped power allocation is a hybrid of the first two combining their benefits.

Equal Power Allocation

The goal to create a Gaussian like distribution can be easily achieved by choosing the same amplitude $\alpha_n = \sqrt{\frac{1}{m}}$ for each bit level. On one hand, the resulting overlap of symbols creates a Bernoulli distribution which is indeed Gaussian like if m is not too small (cf. Fig. 2.14), e.g., $m = 2$ leads to a triangular distribution. On the other hand, this mapping is not bijective due to the introduced ambiguity which makes uncoded transmission impossible. The mapping itself can be considered as a compression which is intuitive because a nonuniform distribution cannot have maximum entropy. Hence, the stream of bits before the mapping must be redundant. For

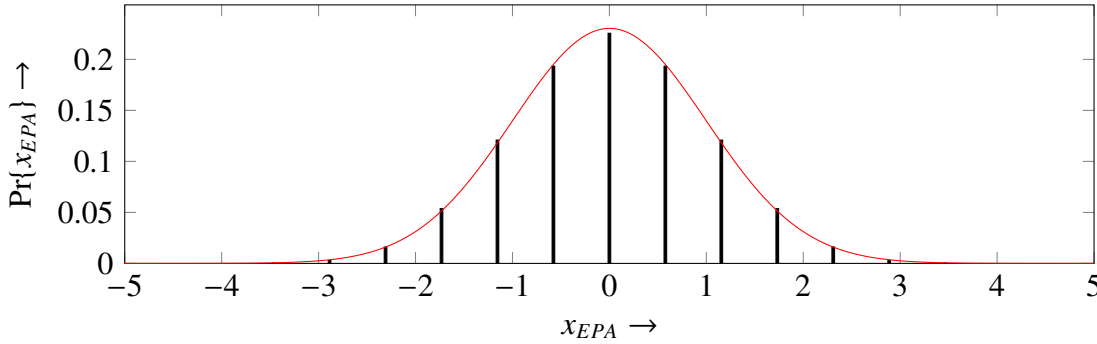


Figure 2.14: Signal space distribution of superposition modulation for equal power allocation with $m = 12$.

equal power allocation, this compression is quite strong which is indeed a drawback because it limits the achievable spectral efficiency $I(X; Y)$ for a given number of bit levels. More precisely, the limiting factor is the entropy $H(X)$ which grows only logarithmically with the number of bit levels m [Wo11].

$$H(X) \approx \frac{1}{2} \log \left(\frac{\pi}{2} em \right) \quad (2.8)$$

For the examples in Fig. 2.14 and Fig. 2.15 with $m = 12$ and $m = 18$ layers, the spectral efficiency $I(X; Y)$ cannot exceed $H(X) = 2.8385$ and $H(X) = 3.1321$ bits per channel use (cf. (2.8)) whereas a bijective mapping would have the respective maximum entropy of $H(X) = 12$ and $H(X) = 18$ bits, respectively. Entropy depends very much on the cardinality $|\mathcal{X}| = m + 1$ which grows only linearly in m . In theory, the strong compression (logarithmic growth of entropy m) is not an issue because the number of bit levels can be increased arbitrarily. In practice however, a higher number of bit levels corresponds with a higher demapping complexity $O(m)$

which grows linear with m for equal power allocation [Wo11, HW11]. Hence, a trade-off arises between complexity and input entropy which should not be lower than the channel capacity.

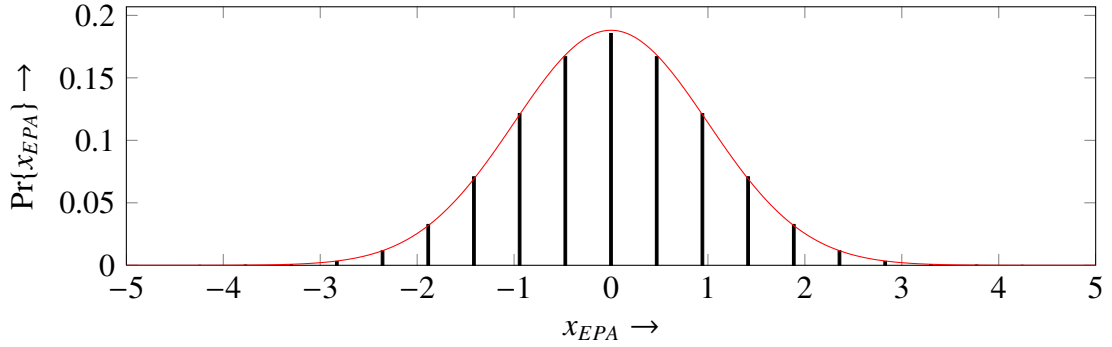


Figure 2.15: Signal space distribution of superposition modulation for equal power allocation with $m = 18$.

Unequal Power Allocation

Based on its name, unequal power allocation could be anything which is not equal. Anyhow, according to [Wo11], α_n is meant to decrease exponentially with n :

$$\alpha_n = \alpha_1 \cdot \xi^{n-1} \quad 0 < \xi < 1 \quad (2.9)$$

where ξ is the exponential base and α_1 is chosen according to an average power constraint. Contrary to equal power allocation, which is nonbijective and follows a nonuniform distribution, unequal power allocation always creates a bijective mapping which follows a uniform distribution as shown in Fig. 2.16. Hence, unequal power allocation has maximum input entropy (no

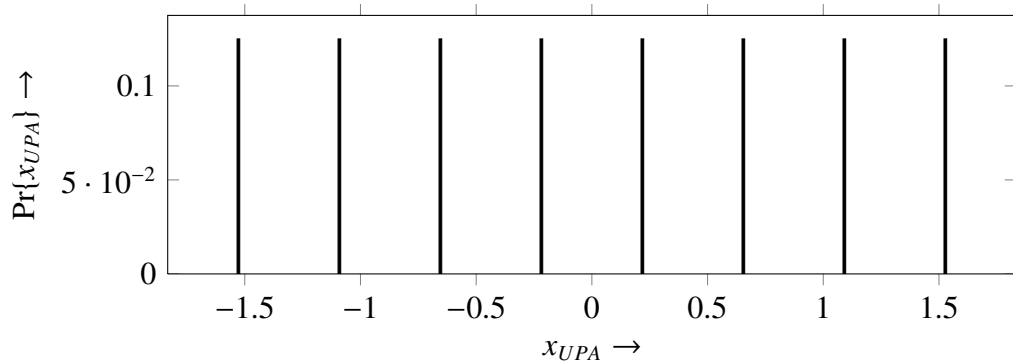


Figure 2.16: Signal space distribution of superposition modulation for unequal power allocation with $m = 3$.

compression) which allows a higher spectral efficiency $I(X; Y)$ compared to equal power allo-

cation with the same number of bit levels.³ On the downside, without compression the shaping gain vanishes too which makes unequal power allocation the other extreme compared to equal power allocation which has a maximum shaping gain and a strong compression.

As illustrated in Fig. 2.16, $\xi = 0.5$ will lead to an equispaced signal space and, thus, equals an ASK with natural labeling. Although, other choices of ξ are possible, a nonequispaced signal space is mostly not desirable and will therefore not be considered in this work. For details, the reader is referred to [Wo11]. The demapping complexity for unequal power allocation is $O(2^m/m)$ [Wo11, HW11].⁴

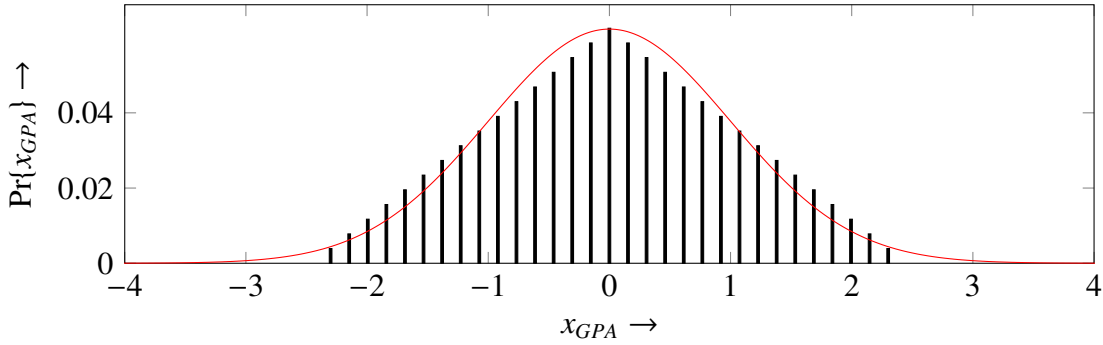


Figure 2.17: Signal space distribution of superposition modulation for grouped power allocation with $m = G \cdot L = 2 \cdot 4 = 8$ and $H(X) = 4.7159$.

Grouped Power Allocation

As shown above, both equal and unequal power allocation have opposed benefits and drawbacks which can be traded off by a hybrid scheme. The goal is to create a Gaussian like distributed signal constellation which at the same time has high input entropy. The key insights from the above schemes are that equal power allocation helps to achieve a Gaussian like shape whereas unequal power allocation helps to increase the cardinality $|\mathcal{X}|$ and, thus, entropy $H(X)$.

Hence, a hybrid scheme can be derived by dividing m layers into L groups of size G whereby levels within a group have the same power while power allocation along the groups is unequal (exponential) with $\xi = 0.5$ [Wo11].

$$x_{GPA} = \sum_{l=1}^L \alpha_l \sum_{g=1}^G c_{l,g} \quad c_{l,g} \in \{-1, +1\}, \alpha_l = \alpha_1 \cdot 2^{1-l} \quad (2.10)$$

As the sums are interchangeable, one can think as well of a sum of unequal power allocation schemes which corresponds to a discrete convolution of uniform (rectangular) distributions

³Contrarily to equal power allocation, the entropy $H(X) = m$ of unequal power allocation grows linearly with m instead of logarithmically, that is, cardinality $|\mathcal{X}| = 2^m$ grows exponentially instead of linearly.

⁴As this scheme is comparable to ASK, one may expect a complexity of $O(2^m)$. That is true for Gray labeling but not for natural labeling inherent in superposition modulation.

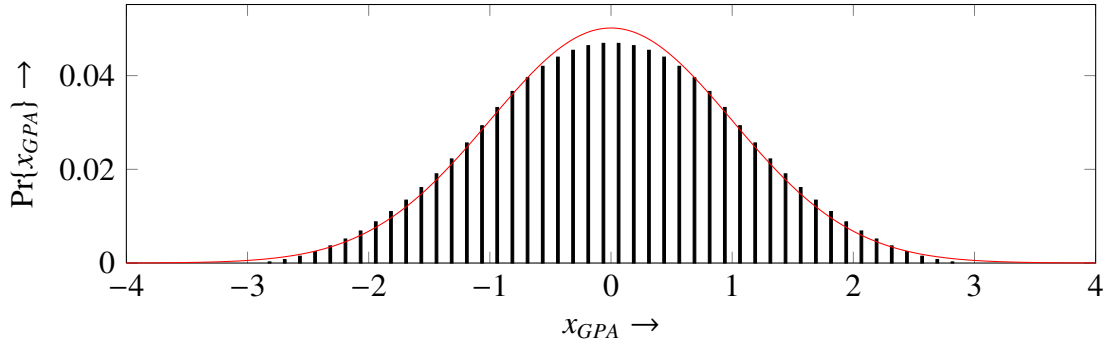


Figure 2.18: Signal space distribution of superposition modulation for grouped power allocation with $m = G \cdot L = 3 \cdot 4 = 12$ $H(X) = 5.0345$.

(cf. Fig. 2.17, Fig. 2.18, Fig. 2.19). Consequently, $G = 1$ directly leads to unequal power allocation, $G = 2$ leads to a triangular distribution (cf. Fig. 2.17), and $G = 3$ already achieves a very good Gaussian like shape (cf. Fig. 2.18). Larger values of $G > 3$ further improve the shaping as shown in Fig. 2.19 until it is equal to the equal power allocation scheme with $G = m$. However, this does not have much impact on the shaping gain which is the gap to Shannon's

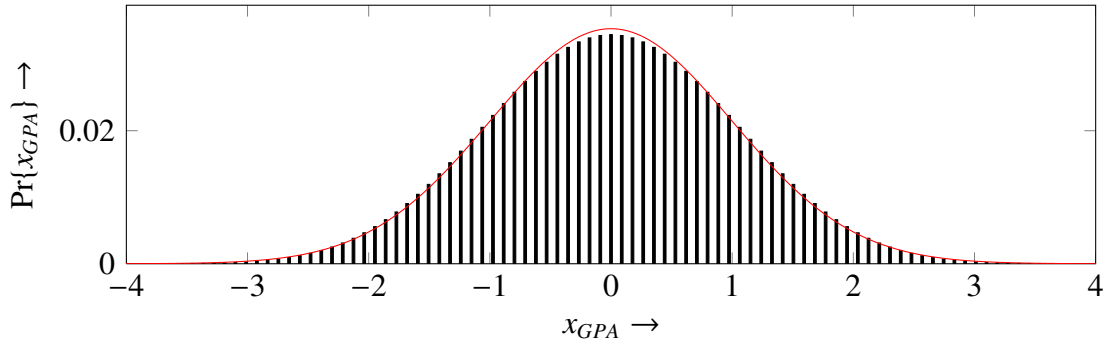


Figure 2.19: Signal space distribution of superposition modulation for grouped power allocation with $m = G \cdot L = 6 \cdot 4 = 18$ and $H(X) = 5.0724$.

capacity: The achievable rates of different superposition modulation schemes are depicted in in Fig. 2.20 where different colors indicate different power allocations, different markers belong to different group sizes G and different line styles correspond to the number of levels L or m . It is important to note that the saturation level can be increased by increasing either m for equal power allocation or L for grouped power allocation. Comparing now different grouped power allocation schemes in the non-saturated ranges, a difference between different group sizes G is hardly visible which makes $G = 2$ and $G = 3$ the most interesting group sizes for practical implementations. By comparing the spectral efficiency of equal power allocation and grouped power allocation in terms of mutual information, one can see that all grouped power allocated schemes outperform the equal power allocated schemes. This is reasoned in the growth rate of

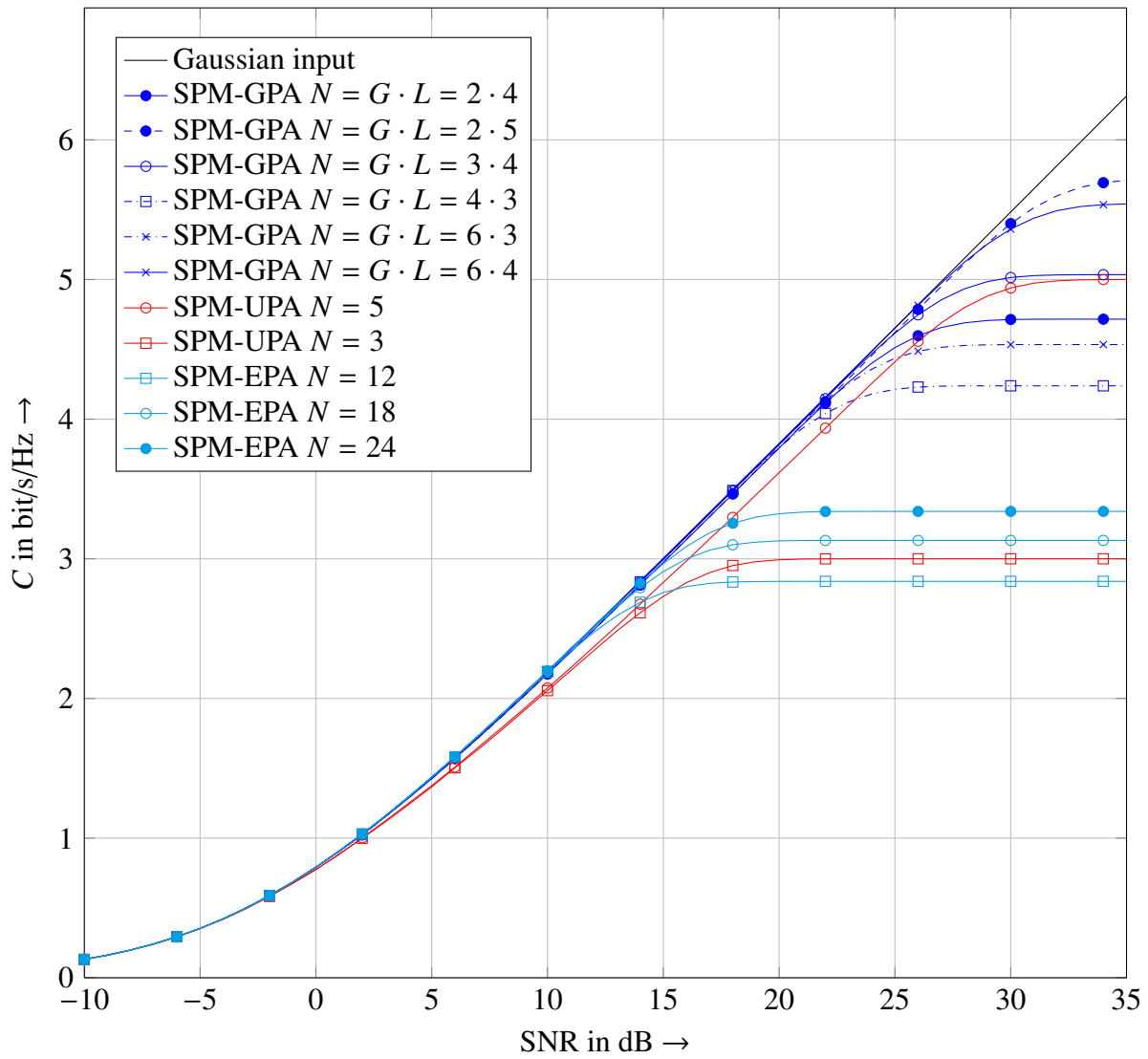


Figure 2.20: Capacity of AWGN channel for superposition modulation with different power allocation schemes.

the entropy with respect to the number of bit levels which limits the achievable mutual information. In particular, the entropy

$$H(X) \approx \frac{1}{2} \log \left(\frac{\pi}{6} eG \right) + L \quad (2.11)$$

for grouped power allocation grows faster than logarithmically (cf. (2.8) for equal power allocation) in the number of layers $m = L \cdot G$. This becomes visible as part of the entropy in (2.11) grows linearly, like for unequal power allocation, while the other part grows logarithmically with the number of layers. Likewise this trade-off is obvious in the cardinality $|\mathcal{X}| = G(2^L - 1) + 1$ which grows linearly with group size G as for equal power allocation and exponentially with the number of power levels L as for unequal power allocation. Thus, for the same number of bit levels m , grouped power allocation delivers higher cardinality than equal power allocation as illustrated in Figures 2.14 and 2.18. Demapping complexity is traded off as well to order $O(G \cdot 2^L/L)$ [Wo11, HW11].

2.3.5 Bit Interleaved Coded Modulation

In contrary to classical multilevel coding which uses a separate code on each bit level, a single code over all bit levels can also achieve the sum capacity of all bit levels. As the bit levels are not independent as the decomposition in (2.6) shows, feedback between demapper and decoder is required which results in an iterative turbo processing similar to that depicted in Fig. 2.7. Especially when using turbo coding, such an iterative processing can become very complex. To avoid the turbo loop between demapper and decoder, one could neglect the conditions in (2.6) which leads to a truly parallel transmission of bits. On one hand, the performance must be worse because the available information is not fully exploited. On the other hand, this loss is quite small over a large range of SNRs when Gray mapping is used which makes this scheme very attractive for practical applications. Fig. 2.21 shows the blockdiagram of this BICM scheme with parallel decoding as proposed by Stiersdorfer in [Sti09]. The interleaver is necessary to split burst errors arising from higher order modulation.

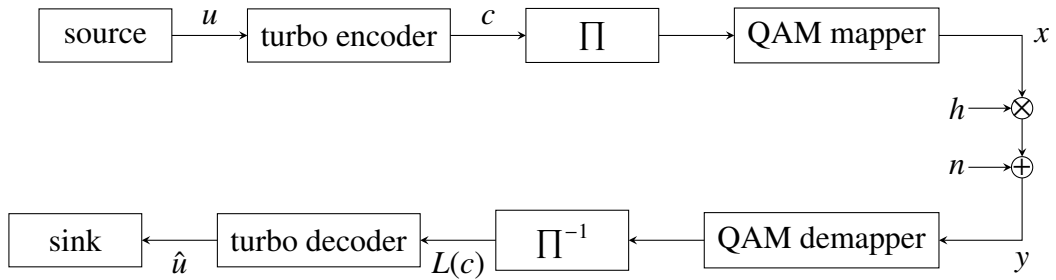


Figure 2.21: Blockdiagram of Bit Interleaved Coded Modulation.

2.4 Simulation Setup and Achievable Rates

For the simulations throughout this work, the BICM scheme (cf. Sec. 2.3.5) is used combining turbo coding (cf. Sec. 2.3.2) with code rates $R_c \in \{1/3, 4/11, 2/5, 4/9, 1/2, 4/7, 2/3, 4/5\}$ and QAM (cf. Sec. 2.3.1) with orders $m \in \{2, 4, 6, 8, 10\}$. The arising 40 MCSs with rates $R_b = m \cdot R_c$ cover the practically relevant range and are all simulated over AWGN channels to obtain Bit Error Rates (BERs) versus SNR. Fig. 2.22 shows achievable rates R_b for a target BER of 10^{-5} versus the respective threshold SNR where different colors belong to different modulation orders m (R_c varied within same color). The Monte Carlo simulation was conducted with an interleaver

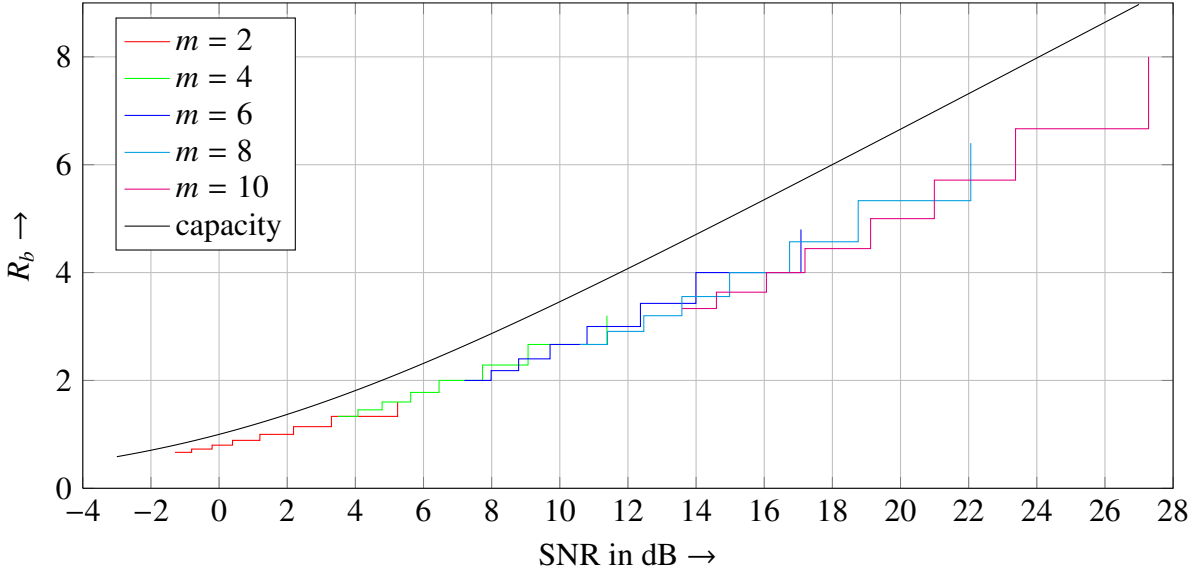


Figure 2.22: Achievable rates R_b versus SNR for $\text{BER} = 10^{-5}$ for AWGN channel.

length of 8640, 8 turbo iterations and 4000 statistical realizations whereby feedback between demapper and decoder was neglected as mentioned in Sec. 2.3.5. The results of Fig. 2.22 will be used in Ch. 5, Ch. 6, and Ch. 7 to determine rates for P2P transmissions occurring within the relay network, e.g., from source to relay, source to destination, or relay to destination.

2.5 Orthogonal Frequency Division Multiplexing

As this work investigates practical issues, OFDM is of great interest since it is the state-of-the-art scheme to deal with frequency selective channels occurring with the use of high bandwidth. The key idea is to subdivide system bandwidth into a bunch of subcarriers of small bandwidth so that the channel transfer function in each subband is almost perfectly flat [Kam08]. Orthogonality is achieved by means of rectangular shaping in time domain which corresponds to a sum of sines in frequency domain. Hence, with appropriate carrier spacing Δf , the 1. Nyquist criterion is fulfilled in both time and frequency domain. Fig. 2.23 shows the overlapping subcarriers

before summation to visualize that for exact sampling only one subcarrier is visible while all others have a zero. The mixing of data symbols to the different subcarriers can be easily imple-

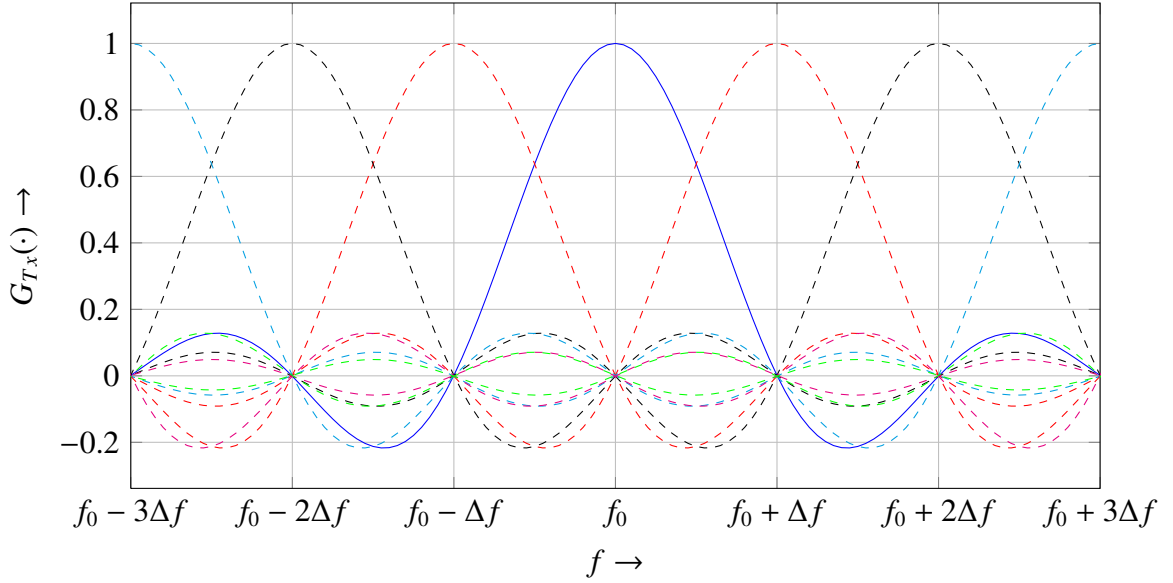


Figure 2.23: Sinc spectrum of each subcarrier $G_{Tx}(\cdot)$ causes no ICI with appropriate carrier spacing Δf around carrier frequency f_0 .

mented by means of the Fast Fourier Transform (FFT) which inherently provides the subcarrier frequencies - one of the reasons why OFDM is very suitable for practical applications. Another important reason is the low complex channel equalization as later shown in Sec. 2.5.2. Usually, benefits in implementation come with the cost of low efficiency: The drawbacks of OFDM are mainly a lower spectral efficiency due to the required guard interval and unused carriers at the spectrum border to combat out-of-band radiation. Furthermore, the superposition of all subcarriers in time domain leads to a high Peak to Average Power Ratio (PAPR) which makes costlier amplifiers necessary.

2.5.1 OFDM - System Setup

The OFDM system as shown in Fig. 2.24 uses the same BICM as described above. Therefore, the signal processing at the source and at the receiver are similar to a single-carrier system except for a Serial-to-Parallel (S/P) converter before the mapper to ensure modulation along different subcarriers. The blocks Inverse Fast Fourier Transform (IFFT) and FFT are then used to assign the data symbols to the OFDM subcarriers and vice versa whereby usually some subcarriers are non-occupied (out-of-band radiation, filter design issues) or used for reference symbols (synchronization, channel training). Thereafter, a guard interval (GI) is added in form of a cyclic prefix to combat Inter-Symbol-Interference (ISI) and Inter-Carrier-Interference (ICI) before the time domain signal is transmitted over the channel. Anyway, for the results in Ch. 7, the sys-

tem in Fig. 2.24 is simulated completely in frequency domain as indicated by the box which summarizes parts of the signal processing and the channel impulse response \mathbf{h}_{TD} into a channel transfer function \mathbf{h} . In particular, \mathbf{h}_{TD} is modeled as a tapped delay line according to sampling at symbol rate so that $\mathbf{h} = [h_1, h_2, \dots, h_{N_{FFT}}]$ contains all channel coefficients (cf. Sec. 2.1.2) of the OFDM subcarriers. In a nutshell, the BICM in Fig. 2.21 is simply extended for all subcarriers.

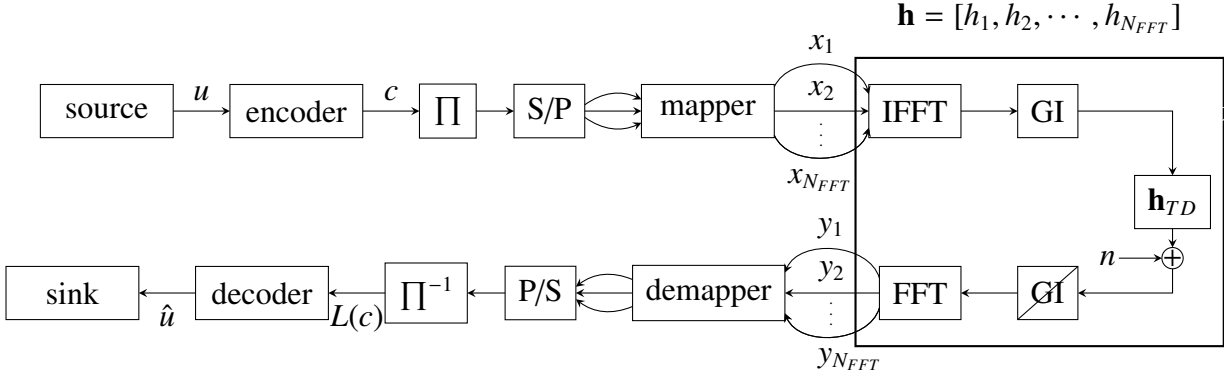


Figure 2.24: Blockdiagram of Orthogonal Frequency Division Multiplexing.

2.5.2 Channel Estimation and Equalization

For channel estimation a pilot symbol x^p is transmitted on each carrier prior to actual data transmission. Dropping the subcarrier index, the destination observes

$$y^p = h \sqrt{P} x^p + n^p \quad (2.12)$$

where $n^p \sim \mathcal{CN}(0, 1)$ is AWGN as in Sec. 2.1. For the sake of clarity, $x^p = 1$ will be assumed to simplify the mathematical description of occurring probability mass functions (pmfs) like $\Pr\{x, \bar{y}, \hat{h}\}$ in Ch. 7. Nonetheless, all derived quantization concepts in Ch. 7 will work in general for arbitrary pilots which is preferable in practice to avoid a high PAPR. Due to the fact that the channel transfer function can be represented by one channel coefficient per subcarrier, channel estimation and equalization become simple scalar operations. The following mathematical description of channel estimation and equalization is written such that the joint pmfs $\Pr\{x, \bar{y}, \hat{h}\}$ of

relevant signal x , equalized observation \bar{y} , and estimated channel coefficient \hat{h} can be derived straightforwardly.⁵

$$\hat{h} = \frac{y^p}{\sqrt{P}x^p} = h + \frac{n^p}{\sqrt{P}} \quad (2.13a)$$

$$\begin{aligned} \bar{y} &= y \cdot \hat{h}^{-1} \\ &= \sqrt{P} \frac{hx}{\hat{h}} + \frac{n}{\hat{h}} \\ &\stackrel{(a)}{=} \sqrt{P} \cdot \frac{h + \frac{n^p}{\sqrt{P}} - \frac{n^p}{\sqrt{P}}}{h + \frac{n^p}{\sqrt{P}}} x + \frac{n}{\hat{h}} \\ &= \sqrt{P} \cdot \left(1 - \frac{\frac{n^p}{\sqrt{P}}}{h + \frac{n^p}{\sqrt{P}}} \right) x + \frac{n}{\hat{h}} \\ &= \sqrt{P} \cdot x - \frac{n^p}{\hat{h}} x + \frac{n}{\hat{h}} \\ &= \sqrt{P} \cdot x - \frac{\bar{n}^p}{|\hat{h}|} x + \frac{\bar{n}}{|\hat{h}|} \end{aligned} \quad (2.13b)$$

In (a), noise term $\frac{n^p}{\sqrt{P}}$ is added and subtracted to get rid off variable h and keep only \hat{h} which is desirable to determine $\Pr\{x, \bar{y}, \hat{h}\}$. In the final step, the channel phases are incorporated into the noise variables $\bar{n} = n \cdot e^{-\angle \hat{h}}$ and $\bar{n}^p = n^p \cdot e^{-\angle \hat{h}}$ since a phase rotation does not change the statistics of circular symmetric AWGN: $n, \bar{n}, n^p, \bar{n}^p \sim \mathcal{CN}(0, 1)$. The channel estimation error is denoted by the term $\frac{\bar{n}^p}{|\hat{h}|} x$ in (2.13b). Whenever perfect channel knowledge is considered, (2.13b) simplifies to

$$\bar{y} = \sqrt{P}x + \bar{n} \cdot |h|^{-1}. \quad (2.13c)$$

2.6 Discussion

Basic concepts of communication and information theory have been discussed for a P2P transmission. The introduced channel model is used throughout the work to model partial links of the relay channel. Furthermore, the channel capacity according to Shannon [Sha48] can be generalized to a more complex network using the Max-Flow Min-Cut (MFMC) theorem [FF56, FF62, CT91, EK12] (cf. Ch. 3). The described MCSs will be used to determine achievable rates for different relaying strategies in a more practical context: In Ch. 4 multilevel coding will be combined with superposition modulation to approach the theoretically achievable rate of DF. Ch. 5 and Ch. 6 contain simulation results for the described BICM scheme. Frequency selective Rayleigh Fading channels with OFDM will only be considered for CF in Ch. 7 because

⁵This classification is used by the Information Bottleneck (IB) method utilized in Ch. 6 and Ch. 7.

they heavily affect quantizer design. The rest of the work considers only AWGN and path loss (cf. Sec. 2.1.1).

Chapter 3

Classical Relay Channel - Preliminaries

This chapter introduces the classical 3-node relay channel and the most important relay protocols: Decode-and-Forward (DF), Compress-and-Forward (CF), and Amplify-and-Forward (AF). After presentation of the 3-node relay channel in Sec. 3.1, an upper bound on the capacity [CE79, KGG05, EK12] is derived in Sec. 3.2 using the Max-Flow Min-Cut (MFMC) theorem of Ford and Fulkerson [FF56, FF62, CT91, EK12]. Sec. 3.3 and Sec. 3.4 introduce the relay protocols DF, CF, and AF as well as their achievable rates for half duplex relays whereby the latter section provides closed form solutions for the well known Gaussian relay channel. Finally, Sec. 3.5 shows and discusses corresponding results for a per node power, a per node energy, and a sum energy constraint. While most of these results are well known in literature [KGG05, HMZ05, HM02, Wei12], one further publication considering transmit energy instead of power per node emerged during the work on this thesis [KK15a].

3.1 System Setup – 3-Node Relay Channel

The system under investigation in this thesis is the memoryless 3-node relay channel as depicted in Fig. 3.1: A source node S wants to transmit a message to destination node D where relay node R supports this transmission. The three emerging partial links, $S \rightarrow R$, $S \rightarrow D$, and $R \rightarrow D$ are modeled as described in Sec. 2.1 whereby $h_{ef} \forall e \in \{S, R\}, \forall f \in \{R, D\}, e \neq f$ denotes generally the channel coefficient depending on distance d_{ef} . The signals X_S and Y_D with realizations $x_S \in \mathcal{X}_S$ and $y_D \in \mathcal{Y}_D$ denote in- and output of the channel, respectively, whereby Y_R and X_R with realizations $y_R \in \mathcal{Y}_R$ and $x_R \in \mathcal{X}_R$ are intermediate receive and transmit signals at the relay. The channel transition probability is denoted by $p(y_R, y_D | x_S, x_R)$. The following general derivations hold for known channel coefficients and general input distributions $p(x_S, x_R)$ with restriction on the average signal power. Sec. 3.4 presents solutions for the Gaussian relay channel for which a Gaussian input is optimal.

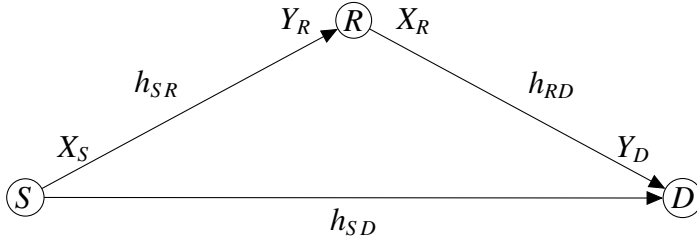


Figure 3.1: 3-node relay Channel.

3.2 Capacity Upper Bound

For a general network with T nodes comprised in set \mathcal{T} and a joint probability density function (pdf) $p(x_1, x_2, \dots, x_T)$, the capacity is upper bounded by [EK12, CT91, FF56, FF62]

$$\max_{p(x_1, \dots, x_T)} \min_{\mathcal{S} \subseteq \mathcal{T}} I(X_{\mathcal{S}}; Y_{\bar{\mathcal{S}}}|X_{\bar{\mathcal{S}}}) \quad (3.1)$$

where \mathcal{T} is divided in two complementary sets \mathcal{S} and $\bar{\mathcal{S}}$ containing transmitting and receiving nodes, respectively. The minimum is taken over the mutual information terms of all possible combinations of transmitting nodes \mathcal{S} to the respective receiving nodes $\bar{\mathcal{S}}$.

3.2.1 Full-Duplex Relays

Applying (3.1) to the 3-node relay channel, as depicted in Fig. 3.2, delivers [EK12, CE79, KGG05]

$$C_{FD} \leq \max_{p(x_S, x_R)} \min\{I(X_S; Y_R, Y_D|X_R), I(X_S, X_R; Y_D)\} \quad (3.2)$$

where $p(x_S, x_R)$ is the joint probability distribution of source's and relay's transmit signals. The two cut-sets can be interpreted as cooperative broadcast and Multiple Access (MAC) bound, i.e., an upper bound is obtained if all transmitting nodes and all receiving nodes cooperate.

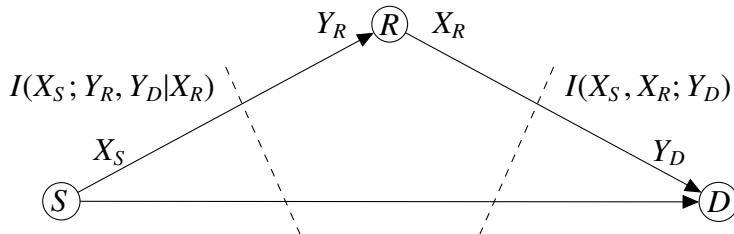


Figure 3.2: Cuts for 3-node relay channel.

3.2.2 Half-Duplex Relays

In the half-duplex case, cooperative MAC and broadcast bounds need to be adapted to two time slots whereby the first time slot will be denoted as broadcast phase and the second as MAC phase. As the relay can either receive or transmit, the respective transmit or receive signals are set to zero. The cooperative broadcast bound simplifies to

$$I(X_S; Y_R, Y_D | X_R) = \begin{cases} I(X_{S1}; Y_R, Y_{D1} | X_R = 0) & \text{for } T_1 \\ I(X_{S2}; Y_R = 0, Y_{D2} | X_R) & \text{for } T_2 \end{cases} \quad (3.3)$$

and the cooperative MAC bound to

$$I(X_S, X_R; Y_D) = \begin{cases} I(X_{S1}, X_R = 0; Y_{D1}) & \text{for } T_1 \\ I(X_{S2}, X_R; Y_{D2}) & \text{for } T_2 \end{cases} \quad (3.4)$$

where subscript numbers denote the corresponding time slots with lengths T_1 and T_2 .¹ The mutual information for each time slot has to be summed for each respective subset whereby the fractional length of the time slots is considered by parameter $\tau \in [0, 1]$ such that $T_1 = \tau T$ and $T_2 = (1 - \tau)T$ [HMZ05].

$$C_{HD}^{non} \leq \max_{\substack{p_{X_{S1}, X_{S2}, X_R} \\ 0 \leq \tau \leq 1}} \min \left\{ \begin{array}{l} \tau I(X_{S1}; Y_R, Y_{D1}) + (1 - \tau) I(X_{S2}; Y_{D2} | X_R) \\ \tau I(X_{S1}; Y_{D1}) + (1 - \tau) I(X_{S2}, X_R; Y_{D2}) \end{array} \right\} \quad (3.5)$$

This capacity bound holds for non-orthogonal channel access since both source and relay are transmitting in the second time slot. Additionally, orthogonal channel access, where the source is quiet in the second time slot, that is, $X_{S2} = 0$, will be considered as well. In this case, (3.5) simplifies to

$$C_{HD}^{orth} \leq \max_{\substack{p_{X_{S1}, X_{S2}, X_R} \\ 0 \leq \tau \leq 1}} \min \left\{ \begin{array}{l} \tau I(X_{S1}; Y_R, Y_{D1}) \\ \tau I(X_{S1}; Y_{D1}) + (1 - \tau) I(X_R; Y_{D2}) \end{array} \right\}. \quad (3.6)$$

3.3 Relaying Strategies and Achievable Rates

The most common relay protocols in the literature are

- Amplify-and-Forward (AF) where the relay simply amplifies its received signal before forwarding it to the destination (time slots of equal length are required),

¹Relay signals are not distinguished because it is obvious that y_R is received in the first and x_R transmitted in the second time slot due to the half-duplex constraint.

- Decode-and-Forward (DF) where the relay must fully decode its received signal and forwards a re-encoded signal (time slot length can be optimized),
- Compress-and-Forward (CF) where the relay cannot decode the message and, thus, forwards only a compressed version of its observed signal (time slot length can be optimized).

Please note that DF refers to exploiting incremental redundancy. For details on DF with repetition coding, the reader is referred to [Lan02, LTW04, Wei12]. For the rest of this work, only half-duplex relays are considered due to their higher practical relevance. Details on full-duplex relays can be found in [CE79, KGG05, EK12] whereby the results therein can be adapted to half-duplex relays the same way as in Sec. 3.2.2.

3.3.1 Amplify and Forward

The AF relay protocol is the simplest known since the relay has very low complexity and does not need any digital processing at all [Lan02, LTW04, EK12, Wei12]. The source encodes a message u_1 into x_{S1} and broadcasts it in the first time slot. The relay simply takes its received signal

$$y_R = h_{SR} \sqrt{P_{S1}} x_{S1} + n_R \quad (3.7a)$$

and retransmits

$$x_R = \eta \cdot y_R = \eta \cdot \sqrt{P_{S1}} h_{SR} x_{S1} + \eta \cdot n_R \quad (3.7b)$$

where

$$\eta \leq \sqrt{\frac{P_R}{P_{S1} h_{SR}^2 + 1}} \quad (3.7c)$$

denotes the amplification which is restricted by source's and relay's transmit power P_{S1} and P_R in first and second time slot, respectively. The respective received signals at the destination in first and second time slot are

$$y_{D1} = h_{SD} \sqrt{P_{S1}} x_{S1} + n_{D1} \quad (3.7d)$$

$$\begin{aligned} y_{D2}^{non} &= h_{SD} \sqrt{P_{S2}} x_{S2} + h_{RD} x_R + n_{D2} \\ &= h_{SD} \sqrt{P_{S2}} x_{S2} + h_{RD} \eta h_{SR} \sqrt{P_{S1}} x_{S1} + h_{RD} \eta n_R + n_{D2} \end{aligned} \quad (3.7e)$$

$$y_{D2}^{orth} = h_{RD} \eta h_{SR} \sqrt{P_{S1}} x_{S1} + h_{RD} \eta n_R + n_{D2} \quad (3.7f)$$

where y_{D2} is distinguished with respect to orthogonal and non-orthogonal MAC. The signal x_{S2} is transmitted additionally (superposition coding) by the source in the second time slot and corresponds to an independently low rate encoded message u_2 . Hence, y_{D2}^{non} is a superposition of the transmit signals x_{S2} and x_R .

By inspection of y_{D1} and y_{D2} it becomes obvious that both observations contain the indirect transmit signal x_{S1} . Hence, the SNR can be optimally increased via maximum ratio combining requiring equally long time slots. In the case of non-orthogonal channel access, successive interference cancellation is applied to first decode x_{S1} from y_{D2}^{non} by treating interference from x_{S2} as noise and subsequently decode x_{S2} after subtracting the influence of x_{S1} . From information theory, the achievable rate with time sharing parameter $\tau = 0.5$ is [Wei12]

$$R_{AF}^{non} = 0.5 \cdot I(X_{S1}; Y_{D1}, Y_{D2}^{non}) + 0.5 \cdot I(X_{S2}; Y_{D2}^{non} | X_{S1}), \quad (3.8)$$

where the first term denotes the transmission of x_{S1} and the second term the additional direct transmission of x_{S2} . The condition in $I(X_{S2}; Y_{D2}^{non} | X_{S1})$ is due to the fact that x_{S1} has been detected and canceled before x_{S2} is decoded. From an information theoretic point of view, one could consider the total information $0.5 \cdot I(X_{S1}, X_{S2}; Y_{D1}, Y_{D2}^{non})$ that flows from source to destination. Then (3.8) follows directly from the chain rule of mutual information and the fact that Y_{D1} shares no information with X_{S2} . For orthogonal channel access, the second information term in (3.8) vanishes and the first will increase ($I(X_{S1}; Y_{D2}^{orth}) \geq I(X_{S1}; Y_{D2}^{non})$) due to the disappearing interference.

$$R_{AF}^{orth} = 0.5 \cdot I(X_{S1}; Y_{D1}, Y_{D2}^{orth}) \quad (3.9)$$

3.3.2 Decode and Forward

The DF relaying protocol is a good choice when the $S \rightarrow R$ link is not the bottleneck of the system because the relay R must be able to decode the source's transmit signal [Wei12, CE79, EK12, KGG05]. Hence, DF is usually applied when the relay is placed closer to the source than to the destination. Similar as before, the source has two messages u_1 and u_2 which are encoded into x_{S1} and \tilde{x}_{S2} whereby the latter signal differs from prior notation due to a different transmission strategy in the MAC phase. Firstly, the source will broadcast x_{S1} which is overheard by relay and destination.

$$y_R = h_{SR} \cdot \sqrt{P_{S1}} \cdot x_{S1} + n_R \quad (3.10a)$$

$$y_{D1} = h_{SD} \cdot \sqrt{P_{S1}} \cdot x_{S1} + n_{D1} \quad (3.10b)$$

The relay decodes u_1 from y_R and re-encodes it to x_R which is forwarded to the destination in the MAC phase. Again it will be distinguished between orthogonal MAC, where the source is quiet, and general non-orthogonal MAC where the source transmits

$$x_{S2} = \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + \rho \cdot x_R \quad (3.11)$$

which is a superposition of x_R and \tilde{x}_{S2} containing u_2 . The parameter

$$\rho = E \{X_{S2} \cdot X_R^*\}$$

denotes the correlation of source's and relay's transmit signals x_{S2} and x_R and determines the amount ρ^2 of transmit power P_{S1} that is used to transmit the same signal x_R as the relay. The residual power $(1 - \rho^2)P_{S1}$ is used to transmit the low rate encoded u_2 via \tilde{x}_{S2} . Hence, the source applies a mixed strategy of distributed beamforming and superposition coding with the extreme cases $\rho = 1$ (only beamforming) and $\rho = 0$ (only superposition coding). Source and relay can be interpreted as a virtual transmitter with two antennas that are either used for beamforming to increase the array gain or for spatial multiplexing. As beamforming means to adjust the carrier phases of the transmit signals such that a constructive superposition is achieved at the destination, the source needs knowledge of its own and the relay's channel to the destination. The optimal strategy, i.e., the optimal value of ρ depends on the setup. Usually, distributed beamforming (large ρ) is preferable when source and relay are very close compared to the destination because the array gain is maximal for equal channel gains. The respective receive signals at the destination for non-orthogonal and orthogonal channel access in the MAC phase are

$$\begin{aligned} y_{D2}^{non} &= h_{SD} \sqrt{P_{S2}} \cdot x_{S2} + h_{RD} \sqrt{P_R} \cdot x_R + n_{D2} \\ &= h_{SD} \cdot \sqrt{P_{S2}} \cdot (\sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + \rho \cdot x_R) + h_{RD} \cdot \sqrt{P_R} \cdot x_R + n_{D2} \\ &= h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + (h_{SD} \sqrt{P_{S2}} \cdot \rho + h_{RD} \sqrt{P_R}) \cdot x_R + n_{D2} \end{aligned} \quad (3.12a)$$

$$y_{D2}^{orth} = h_{RD} \sqrt{P_R} \cdot x_R + n_{D2}. \quad (3.12b)$$

In the case of non-orthogonal channel access, the destination applies successive interference cancellation. Firstly, x_R will be decoded jointly from y_{D1} and y_{D2}^{non} whereby \tilde{x}_{S2} is treated as noise. Then, after subtracting the influence of x_R in (3.12a), \tilde{x}_{S2} will be decoded. Likewise as in Sec. 3.2.2, the achievable rate becomes [CE79, KGG05, EK12]

$$R_{DF}^{non} \leq \max_{\substack{p(x_{S1}, x_{S2}, x_R) \\ 0 \leq \tau \leq 1}} \min \left\{ \begin{array}{l} \tau I(X_{S1}; Y_R) + (1 - \tau) I(X_{S2}; Y_{D2}^{non} | X_R) \\ \tau I(X_{S1}; Y_{D1}) + (1 - \tau) I(X_{S2}, X_R; Y_{D2}^{non}) \end{array} \right\} \quad (3.13)$$

for non-orthogonal MAC and

$$R_{DF}^{orth} \leq \max_{\substack{p(x_{S1}, x_R) \\ 0 \leq \tau \leq 1}} \min \left\{ \begin{array}{c} \tau I(X_{S1}; Y_R) \\ \tau I(X_{S1}; Y_{D1}) + (1 - \tau) I(X_R; Y_{D2}^{orth}) \end{array} \right\}. \quad (3.14)$$

for orthogonal MAC. Both are similar to (3.5) and (3.6) except the restriction $I(X_{S1}; Y_R, Y_{D1}) \rightarrow I(X_{S1}; Y_R)$ since the relay must be able to decode u_1 from y_R .

3.3.3 Compress and Forward

The relay strategy CF is useful when the relay is not able to successfully decode the messages from the source, e.g., in scenarios where the relay is placed considerably closer to the destination than to the source. In such a scenario, the relay can still cooperate by forwarding not the full message but a compressed version \hat{y}_R of its observation which is usually a quantized version of the received signal y_R . As this compression is correlated to the signal the destination received in the first time slot, the relay applies Wyner-Ziv coding [WZ76, Wyn78] to decrease the necessary rate to forward \hat{y}_R . Similar as for DF, the source broadcasts x_{S1} containing u_1 in the first time slot. In the second time slot, the relay transmits x_R which contains only part of the message u_1 . In particular, the relay compresses, e.g., via quantization, its received signal y_R to an index \hat{y}_R which is Wyner-Ziv source encoded via random binning (cf. App. B) into bin index $b \in \mathcal{B}$ [EK12, Zei12, SW73]. Bin index b is channel encoded and mapped to x_R . Again, depending on channel access being orthogonal or non-orthogonal, the source may transmit x_{S2} additionally in the second time slot which contains, contrary to DF, only message u_2 because it cannot know signal x_R the relay will transmit – index \hat{y}_R is noise dependent. The received signals at relay and destination in first and second time slot are

$$y_R = h_{SR} \cdot \sqrt{P_{S1}} \cdot x_{S1} + n_R, \quad (3.15a)$$

$$y_{D1} = h_{SD} \cdot \sqrt{P_{S1}} \cdot x_{S1} + n_{D1}, \quad (3.15b)$$

$$y_{D2}^{non} = h_{SD} \sqrt{P_{S2}} \cdot x_{S2} + h_{RD} \sqrt{P_R} \cdot x_R + n_{D2}, \quad (3.15c)$$

$$y_{D2}^{orth} = h_{RD} \sqrt{P_R} \cdot x_R + n_{D2}. \quad (3.15d)$$

To detect the messages u_1 and u_2 , the destination needs to firstly recover the bin index b via successive interference cancellation, that is, decoding of y_{D2}^{non} with respect to x_R treating x_{S2} as noise. Then, u_2 can be detected after subtracting $h_{RD} \sqrt{P_R} \cdot x_R$ from y_{D2}^{non} . For orthogonal channel access, b is detected from y_{D2}^{orth} . The uncertainty (about \hat{y}_R due to Wyner-Ziv coding) in b is resolved with the help of y_{D1} (cf. App. B) which serves as side information: Finally, \hat{y}_R and y_{D1}

are jointly decoded to obtain u_1 .² The achievable rates according to the MFMC theorem for non-orthogonal and orthogonal channel access are [EK12]

$$R_{CF}^{non} \leq \max_{p(x_{S1})p(x_{S2})p(x_R)p(\hat{y}_R|y_R)} \min \left\{ \begin{aligned} &\tau \cdot I(X_{S1}; \hat{Y}_R, Y_{D1}) + (1 - \tau) \cdot I(X_{S2}; Y_{D2}^{non}|X_R) \\ &\tau \cdot (I(X_{S1}; Y_{D1}) - I(Y_R; \hat{Y}_R|X_{S1}, Y_{D1})) + (1 - \tau) \cdot I(X_R, X_{S2}; Y_{D2}^{non}) \end{aligned} \right\} \quad (3.16)$$

and

$$R_{CF}^{orth} \leq \max_{p(x_{S1})p(x_R)p(\hat{y}_R|y_R)} \min \left\{ \begin{aligned} &\tau \cdot I(X_{S1}; \hat{Y}_R, Y_{D1}) \\ &\tau \cdot (I(X_{S1}; Y_{D1}) - I(Y_R; \hat{Y}_R|X_{S1}, Y_{D1})) + (1 - \tau) \cdot I(X_R; Y_{D2}^{orth}) \end{aligned} \right\}, \quad (3.17)$$

respectively. Similar as in (3.5) and (3.6), the upper cut-set denotes the broadcast bound with Y_R replaced by \hat{Y}_R due to the compression loss while the lower cut-set denotes the MAC bound with subtracted term $\tau \cdot I(Y_R; \hat{Y}_R|X_{S1}, Y_{D1})$ representing the irrelevance remaining in \hat{Y}_R (and thus in X_R) due to noisy Y_R . In particular, $\tau \cdot I(Y_R; \hat{Y}_R|X_{S1}, Y_{D1})$ is a penalty term with respect to $(1 - \tau) \cdot I(X_R, X_{S2}; Y_{D2}^{non})$ in (3.16) or $(1 - \tau) \cdot I(X_R; Y_{D2}^{orth})$ in (3.17) which contain this irrelevant information about the noise at the relay. According to [EK12], (3.16) can be equivalently described as

$$R_{CF}^{non} \leq \max_{p(x_{S1})p(x_{S2})p(x_R)p(\hat{y}_R|y_R)} \tau \cdot I(X_{S1}; \hat{Y}_R, Y_{D1}) + (1 - \tau) \cdot I(X_{S2}; Y_{D2}^{non}|X_R) \quad (3.18)$$

such that

$$\tau \cdot I(Y_R; \hat{Y}_R|Y_{D1}) \leq (1 - \tau) \cdot I(X_R; Y_{D2}^{non}) \quad (3.19)$$

which is the well known result from [CE79, KGG05], solely adapted to a half-duplex relay.³ Similarly, for orthogonal channel access, (3.17) simplifies to

$$R_{CF}^{orth} \leq \max_{p(x_{S1})p(x_R)p(\hat{y}_R|y_R)} \tau \cdot I(X_{S1}; \hat{Y}_R, Y_{D1}) \quad (3.20)$$

such that

$$\tau \cdot I(Y_R; \hat{Y}_R|Y_{D1}) \leq (1 - \tau) \cdot I(X_R; Y_{D2}^{orth}). \quad (3.21)$$

²Actually, this fact is outdated such that both \hat{y}_R and u_1 can be decoded jointly from b and y_{D1} instead of one after another [EK12]. The generalization of this concept to a bigger network is known as noisy network coding [EK12] whose performance is generally superior. For the 3-node relay channel however, both strategies achieve the same rate as shown in (3.16) and (3.18).

³The equivalent characterization in (3.16) and (3.17) is a recent result following the idea of noisy network coding [EK12].

3.4 The Gaussian Relay Channel

In the Gaussian Relay Channel, AWGN is assumed at all relay nodes. As a Gaussian maximizes entropy (cf. Sec. 2.2 and [CT91]) under a given average power constraint, all inputs X are chosen to be Gaussian as well. In this case, the beforehand presented bounds and achievable rates have a closed form solution as shown in the following. For details on the derivation, the reader is referred to [Wei12, EK12, KGG05, HMZ05].

3.4.1 Capacity Upper Bound

For the Gaussian relay channel, (3.5) and (3.6) simplify to [HMZ05, Wei12]

$$C_{hd} \leq \max_{\substack{0 \leq \rho \leq 1 \\ 0 \leq \tau \leq 1}} \min \left\{ \begin{array}{l} \tau \log_2 \left(1 + (|h_{SR}|^2 + |h_{SD}|^2)P_{S1} \right) + (1 - \tau) \log_2 \left(1 + (1 - \rho^2)|h_{SD}|^2P_{S2} \right) \\ \tau \log_2 \left(1 + |h_{SD}|^2P_{S1} \right) + (1 - \tau) \log_2 \left(1 + |h_{SD}|^2P_{S2} + |h_{RD}|^2P_R + 2\rho h_{SD}h_{RD} \sqrt{P_{S2}P_R} \right) \end{array} \right\}, \quad (3.22)$$

and

$$C_{hd} \leq \max_{\substack{0 \leq \rho \leq 1 \\ 0 \leq \tau \leq 1}} \min \left\{ \begin{array}{l} \tau \log_2 \left(1 + (|h_{SR}|^2 + |h_{SD}|^2)P_{S1} \right) \\ \tau \log_2 \left(1 + |h_{SD}|^2P_{S1} \right) + (1 - \tau) \log_2 \left(1 + |h_{RD}|^2P_R \right) \end{array} \right\}, \quad (3.23)$$

respectively.

3.4.2 Amplify and Forward

Due to maximum ratio combining, the first term in (3.8) can easily be determined by (2.3c) whereby the combined SNR is the sum of SNR according to (3.7d) and Signal to Interference plus Noise Ratio (SINR) according to (3.7e).

$$I(X_{S1}; Y_{D1}, Y_{D2}^{non}) = \log_2 \left(1 + P_{S1}|h_{SD}|^2 + \frac{P_{S1}|h_{RD}|^2|h_{SR}|^2\eta^2}{P_{S2}|h_{SD}|^2 + \eta^2|h_{RD}|^2 + 1} \right) \quad (3.24)$$

The second term in (3.8) corresponds to a P2P transmission from S to D after successive interference cancellation, that is, subtracting $h_{RD}\eta h_{SR} \sqrt{P_{S1}}x_{S1}$ from (3.7e).

$$I(X_{S2}; Y_{D2}^{non}|X_{S1}) = \log_2 \left(1 + \frac{P_{S2}|h_{SD}|^2}{\eta^2|h_{RD}|^2 + 1} \right) \quad (3.25)$$

Hence, (3.8) becomes [Wei12, Lan02, LTW04]

$$R_{AF}^{non} = 0.5 \cdot \log_2 \left(1 + P_{S1}|h_{SD}|^2 + \frac{P_{S1}|h_{RD}|^2|h_{SR}|^2\eta^2}{P_{S2}|h_{SD}|^2 + \eta^2|h_{RD}|^2 + 1} \right) + 0.5 \cdot \log_2 \left(1 + \frac{P_{S2}|h_{SD}|^2}{\eta^2|h_{RD}|^2 + 1} \right). \quad (3.26)$$

For orthogonal channel access, (3.9) simplifies to

$$R_{AF}^{orth} = 0.5 \cdot \log_2 \left(1 + P_{S1}|h_{SD}|^2 + \frac{P_{S1}|h_{RD}|^2|h_{SR}|^2\eta^2}{\eta^2|h_{RD}|^2 + 1} \right) \quad (3.27)$$

which is very similar to (3.24) but without the interference term $P_{S2}|h_{SD}|^2$ in the denominator of the last term, that is, the SINR in (3.24) becomes an SNR due to orthogonal channel access.

3.4.3 Decode and Forward

The general achievable rate for DF (3.13) simplifies to [HMZ05, Wei12]

$$R_{DF}^{non} \leq \max_{\substack{0 \leq \rho \leq 1 \\ 0 \leq \tau \leq 1}} \min \left\{ \begin{array}{l} \tau \log_2 (1 + |h_{SR}|^2 P_{S1}) + (1 - \tau) \log_2 (1 + (1 - \rho^2)|h_{SD}|^2 P_{S2}) \\ \tau \log_2 (1 + |h_{SD}|^2 P_{S1}) + (1 - \tau) \log_2 (1 + |h_{SD}|^2 P_{S2} + |h_{RD}|^2 P_R + 2\rho h_{SD} h_{RD} \sqrt{P_{S2} P_R}) \end{array} \right\} \quad (3.28)$$

for non-orthogonal channel access, and (3.14) to

$$R_{DF}^{orth} \leq \max_{\substack{0 \leq \rho \leq 1 \\ 0 \leq \tau \leq 1}} \min \left\{ \begin{array}{l} \tau \log_2 (1 + |h_{SR}|^2 P_{S1}) \\ \tau \log_2 (1 + |h_{SD}|^2 P_{S1}) + (1 - \tau) \log_2 (1 + |h_{RD}|^2 P_R) \end{array} \right\} \quad (3.29)$$

for orthogonal channel access.

3.4.4 Compress and Forward

The achievable rate for non-orthogonal channel access (3.16) becomes [Wei12, HMZ05]

$$R_{CF}^{non} = \max_{0 \leq \tau \leq 1} \left\{ \tau \log_2 \left(1 + P_{S1}|h_{SD}|^2 + \frac{P_{S1}|h_{SR}|^2}{1 + \sigma_q^2} \right) + (1 - \tau) \log_2 (1 + P_{S2}|h_{SD}|^2) \right\} \quad (3.30)$$

with

$$\sigma_q^2 = \frac{1 + P_{S1}|h_{SD}|^2 + P_{S1}|h_{SR}|^2}{(1 + P_{S1}|h_{SD}|^2) \left(\left(1 + \frac{P_R|h_{RD}|^2}{1 + P_{S2}|h_{SD}|^2} \right)^{\frac{1-\tau}{\tau}} - 1 \right)} \quad (3.31)$$

denoting the variance of the quantizer noise which depends on τ . For orthogonal channel access, (3.17) becomes

$$R_{CF}^{orth} = \max_{0 \leq \tau \leq 1} \tau \log_2 \left(1 + P_{S1}|h_{SD}|^2 + \frac{P_{S1}|h_{SR}|^2}{1 + \sigma_q^2} \right) \quad (3.32)$$

with

$$\sigma_q^2 = \frac{1 + P_{S1}|h_{SD}|^2 + P_{S1}|h_{SR}|^2}{(1 + P_{S1}|h_{SD}|^2) \left((1 + P_R|h_{RD}|^2)^{\frac{1-\tau}{\tau}} - 1 \right)}. \quad (3.33)$$

3.5 Achievable Rates for the Gaussian Relay Channel

For the results in this section, the upper bound and achievable rates for different protocols have been evaluated for the Gaussian relay channel as stated in Sec. 3.4. Therefore, the nodes S , R , and D are assumed to be set up on a line with S at position 0, D at position 1, and R somewhere in between at position $d \in [0, 1]$.⁴ For each of the possible relay positions d , the optimal time sharing parameter τ has to be found. As this cannot be determined analytically, it will be done numerically via an exhaustive search over a finite equidistant grid with a resolution of 1000. For the upper bound and DF, optimal correlation ρ can be determined by equating the respective two cut sets in (3.22) and (3.28). An important question is how to set transmit powers P_{S1} , P_{S2} and P_R at the different nodes in different time slots. A very common way in the literature is to use a power constraint per node which may be further simplified such that each node transmits the same power $P_{S1} = P_{S2} = P_R = P$. Corresponding results are shown in Sec. 3.5.1 and in more detail in [Wei12].

On one hand, a per node power constraint is very pragmatic (suitable for practical implementation) and can be straightforward investigated and analyzed. On the other hand, such a power constraint does not ensure a fair comparison with respect to the transmit energy consumption which is especially of interest for battery-powered nodes. Depending on τ , which is optimized for a specific d , the total but also the individually per node consumed energy varies as shown in the following.

$$E_{sum}^{non} = E_{S1} + E_{S2} + E_R \quad (3.34a)$$

$$= \tau_{non} \cdot T \cdot P_{S1} + (1 - \tau_{non}) \cdot T \cdot (P_{S2} + P_R) \quad (3.34b)$$

$$= \tau_{non} \cdot T \cdot P + (1 - \tau_{non}) \cdot T \cdot 2P \quad (3.34c)$$

$$= (2 - \tau_{non}) \cdot T \cdot P \quad (3.34d)$$

⁴One could extend the range such that R may be left of S or right of D [Wei12], however, this will not deliver deeper insights for the purpose of this thesis.

$$E_{sum}^{orth} = E_{S1} + E_R \quad (3.34e)$$

$$= \tau_{orth} \cdot T P_{S1} + (1 - \tau_{orth}) \cdot T \cdot P_R \quad (3.34f)$$

$$= T \cdot P \quad (3.34g)$$

Furthermore, (3.34) shows that both the individual (per node) as well as the sum transmit energy vary differently for non-orthogonal and orthogonal MAC while sum energy for orthogonal MAC is constant (independent of τ_{orth}). To account for a fair comparison with respect to transmit energy, two additional sharing parameters β and δ are introduced as follows:

1. $\beta \in [0, 1]$ is the fraction of the totally available energy that is used by the source, and
2. $\delta \in [0, 1]$ is the fraction of the source's energy βE_{sum} it can use in the first time slot.

Optimizing τ , β , and δ jointly for each d will lead to an individual power allocation for each node and time slot.

$$P_{S1} = \frac{\beta \cdot \delta}{\tau \cdot T} \cdot E_{sum}^{non} \quad P_{S2} = \frac{\beta \cdot (1 - \delta)}{(1 - \tau) \cdot T} \cdot E_{sum}^{non} \quad P_R = \frac{(1 - \beta)}{(1 - \tau) \cdot T} \cdot E_{sum}^{non} \quad (3.35)$$

This optimal power allocation corresponds to a scenario, where source and relay share a common power supply, e.g., in the downlink of a cellular network. To consider a scenario, e.g., the uplink of a cellular network, where source and relay have local batteries, β has to be fixed, e.g., $\beta = 0.5$. Despite this optimal power allocation, uniform power allocation is considered in the results in Sec. 3.5.2. Given the same total energy consumption, the power allocated to each node is as follows.

$$P = \frac{E_{sum}^{non}}{(2 - \tau_{non}) \cdot T} \quad (3.36)$$

Throughout this thesis, $T = 1$ due to normalization of the noise power to $\sigma_N^2 = \frac{N_0}{T} = 1$ with $N_0 = 1$.

3.5.1 Individual Per Node Power Constraint

Following the above stated line setup, Fig. 3.3 shows the achievable rates for AF, DF, CF and the upper bound on capacity versus relay position d for $\alpha = 4$ and $P = 1$. The same results can be found in [Wei12], including different path-loss exponents α , restrictions like fixing τ , and full duplex relaying [KGG05]. The adaption of τ is quite important as one can see in Fig. 3.3. Thereby it is very meaningful that τ follows somehow d since a short distance implies a low path loss and vice versa, i.e., when the relay R is closer to the source S than to the destination D , one should use more time for the transmission $R \rightarrow D$ than for $S \rightarrow R$.

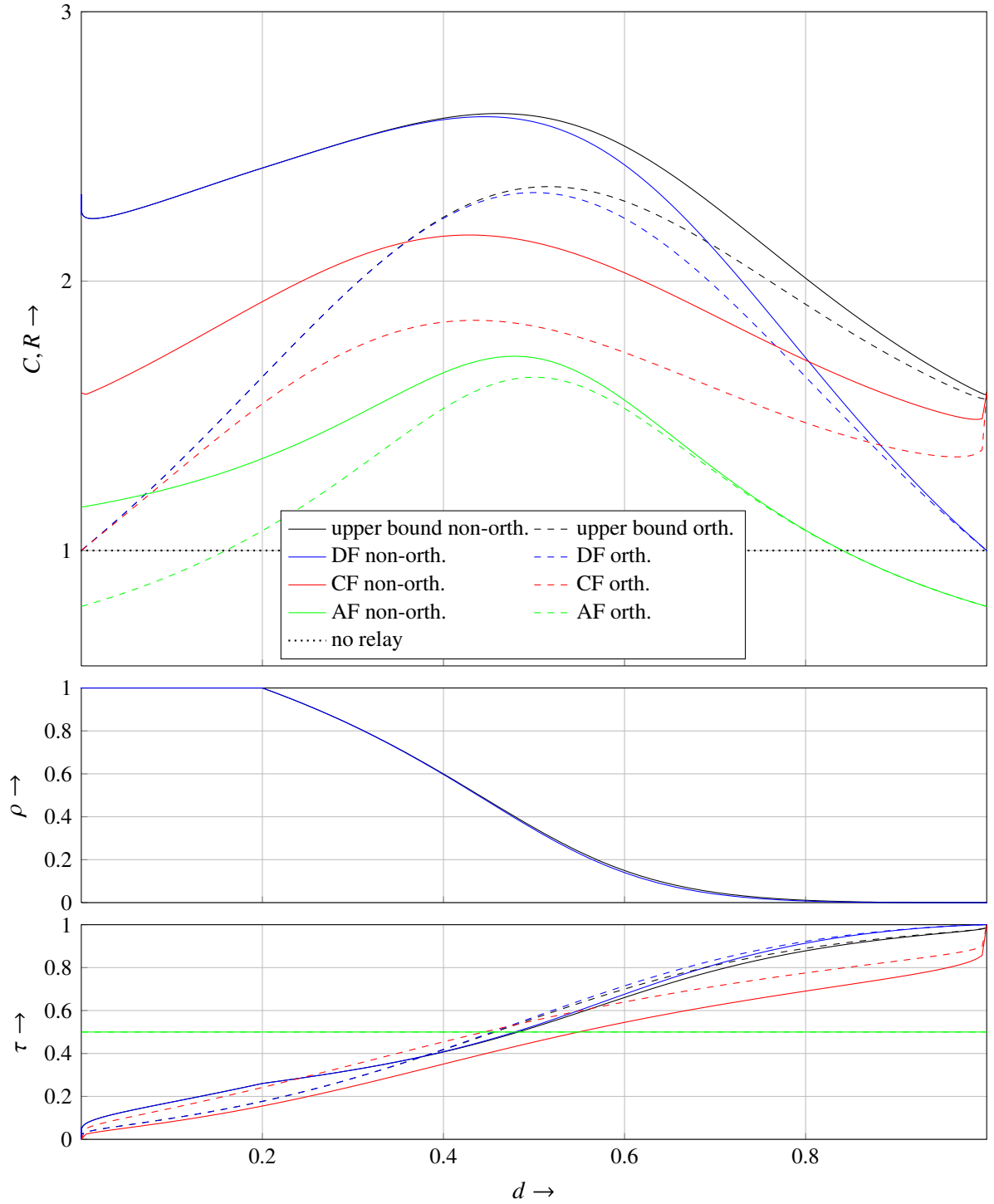


Figure 3.3: Achievable rates, upper capacity bound and optimal parameters ρ and τ versus source relay distance d for individual power constraint with path loss exponent $\alpha = 4$ and transmit power $P = 1$.

Achievable Rates versus Capacity Bound

Let us firstly compare the achievable rates to the upper bound on capacity in Fig. 3.3: For both orthogonal and non-orthogonal channel access, DF operates very close to the capacity for $d < 0.5$, i.e., when the relay is close to the source. For the extreme case $d = 0$, it even reaches the capacity. The more the relay is moved to the destination, the worse DF performs especially for $d > 0.8$ where it gets worse than CF. This is reasoned by the decoding condition at the relay which makes the source-relay link the bottleneck of the system. Moreover, it is also indicated by τ since the DF protocol needs more time than CF in the first time slot especially when the relay is far away from the source. Anyhow, CF cannot close the gap to the capacity bound except for $d = 1$ where it is optimal. Hence, especially for relay positions, where the relay is neither close to the source nor very close to the destination, there is no capacity achieving strategy known so far. The AF relay protocol is always inferior and can even be worse than the direct link due to restriction $\tau = 0.5$. These statements hold for both orthogonal and non-orthogonal MAC whereby even the orthogonal upper bound cannot be reached by non-orthogonal relay protocols in the range about $0.7 < d < 1$.

Non-orthogonal versus Orthogonal Channel Access

Comparing secondly non-orthogonal and orthogonal channel access in Fig. 3.3 reveals huge gains especially for the upper bound and DF in the range $d < 0.4$ where the relay is very close to the source. There are two reasons for this gain: Firstly, owing the individual power constraint per node, the non-orthogonal scheme uses more energy in total and, hence, must be at least as good as the orthogonal scheme plus the gain due to the additional transmit energy. Alternatively, one could think of the virtual direct transmission of \tilde{x}_{s2} in the MAC phase due to superposition coding which gives these additional rate. The power spent for this transmission is determined by the correlation coefficient ρ , precisely the factor $(1 - \rho^2)$, i.e., when $\rho = 0$, all power of the source in the second time slot is used for transmission of \tilde{x}_{s2} . Please note that for CF and AF $\rho = 0$ is mandatory since beamforming is not applicable. This is the reason why the performance gain due to non-orthogonal channel access for CF and AF will (almost) vanish for the sum energy constraint as shown in Sec. 3.5.2. For DF however, there is an additional array gain due to the amount of distributed beamforming for $\rho > 0$ which is maximum in the range $0 \leq d \leq 0.2$, where $\rho = 1$, but also present for higher values of d where $0 < \rho < 1$.

3.5.2 Sum Energy Constraint

The results in this section have been published during the work on this thesis in [KK15a]. Very interesting is the fact that, in contrary to an individual power constraint, the performance gain

due to non-orthogonal relaying is almost fully originated in the virtual array gain of distributed beamforming. The gain due to superposition coding is almost negligible. Firstly, it will be shown how the results of the previous section change when all nodes transmit with the same power P but restricted by sum energy as in (3.36). Secondly, results for optimal power allocation will be shown whereby individual transmit energies per node are not restricted.

Uniform Power Allocation

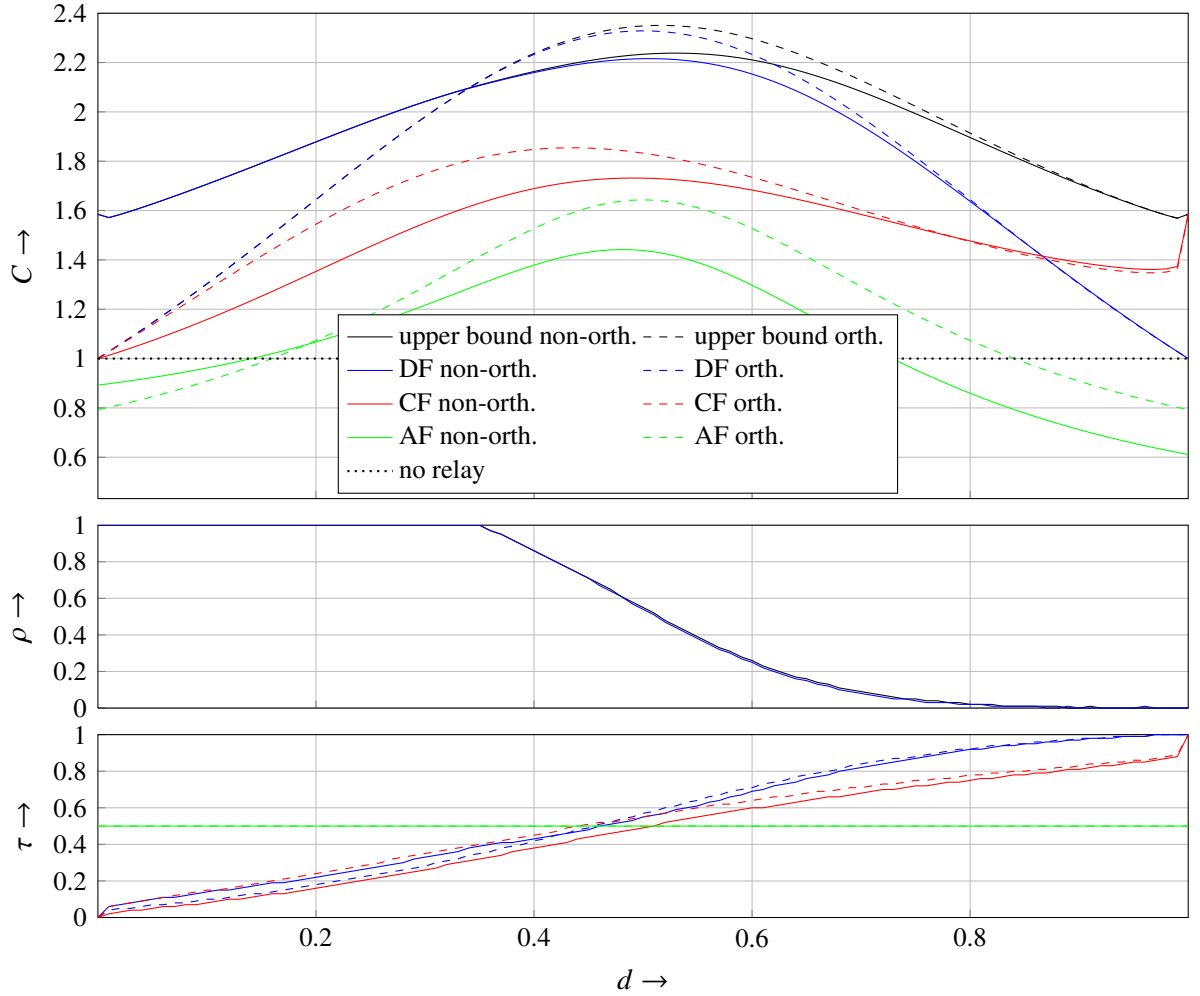


Figure 3.4: Achievable rates, upper capacity bound and optimal parameters ρ and τ versus source relay distance d for uniform power allocation with pathloss exponent $\alpha = 4$ and sum energy $E_{sum}^{non} = 1$.

Fig. 3.4 shows as before the upper bound and achievable rates versus d for a given sum energy with a uniform power allocation. Comparing Fig. 3.4 with Fig. 3.3 reveals that non-orthogonal channel access does not always pay off when considering sum energy instead of individual power. Furthermore, it becomes obvious that the gain of non-orthogonal versus orthogonal MAC is indeed twofold in Fig. 3.3. In Fig. 3.4, there is only a huge gain for DF

due to distributed beamforming when the relay is close to the source. The beforehand gain due to superposition coding in Fig. 3.3 however, vanishes for all DF, CF, and in Fig. 3.4 due to normalization to the sum energy and even becomes a loss since uniform power allocation is not optimal. In the following, this tendency will be confirmed such that the general non-orthogonal scheme is only superior for DF when $\rho = 1$. For $\rho = 0$, DF (AF and CF behave almost in the same manner) breaks down to orthogonal transmission and for $0 < \rho < 1$, the gain is negligible. All other insights from Sec. 3.5.1 still hold in a relative manner.

Optimal Power Allocation

Fig. 3.5 depicts the respective results for AF, DF and CF as well as the upper bound on capacity for an optimal power allocation. Hence, parameters β and δ are plotted too. The comparison of the bound and the protocols with each other leads to the same insights as drawn from Fig. 3.3. However, the respective comparison of non-orthogonal and orthogonal channel access is different: For the upper bound and DF, it becomes obvious that the only gain due to non-orthogonal access is caused by the array gain of distributed beamforming as indicated by ρ which is equal or almost equal to 1 for $d \leq 0.5$. For $d > 0.5$, ρ is distinctly smaller than 1 or equal to 0 while no noticeable gain in the achievable DF rate is visible.⁵ For the first case ($\rho = 1$), the maximum gain is achieved the closer the relay is placed to the source because then most of the source's energy (shared half with relay due to $\beta \approx 0.5$) can be used in the second time slot to apply distributed beamforming, i.e., δ is small and increases with d due to the growing path loss between source and relay. For the latter case ($\rho < 1$), non-orthogonal and orthogonal scheme are almost equal as indicated by $\delta \approx 1$ and a similar transition of τ . Hence, non-orthogonal channel access does only pay off when the relay is close to the source where distributed beamforming is meaningful. A similar behavior is visible when considering CF and AF. For CF, the optimization of the non-orthogonal scheme ends up in the orthogonal except for $d > 0.5$, where a small gain due to superposition in the second time slot ($\delta < 1$) is visible. For AF, optimization delivers the direct transmission scheme for $d < 0.1$ or the orthogonal otherwise. Considering β for all shown rates (for non-orthogonal DF and AF from $d > 0.5$ and $d > 0.1$, respectively), more energy is used by the source with increasing d due to the decreasing path loss on the source-relay link. The behavior of τ is similar to Fig. 3.3.

⁵The slight oscillating of ρ (and also the other parameters) is due to numerical reasons: As the gain is negligibly small, any ρ will lead to almost the same rate as for the orthogonal scheme. Hence, due to limited machine accuracy, the maximum function becomes ambiguous.

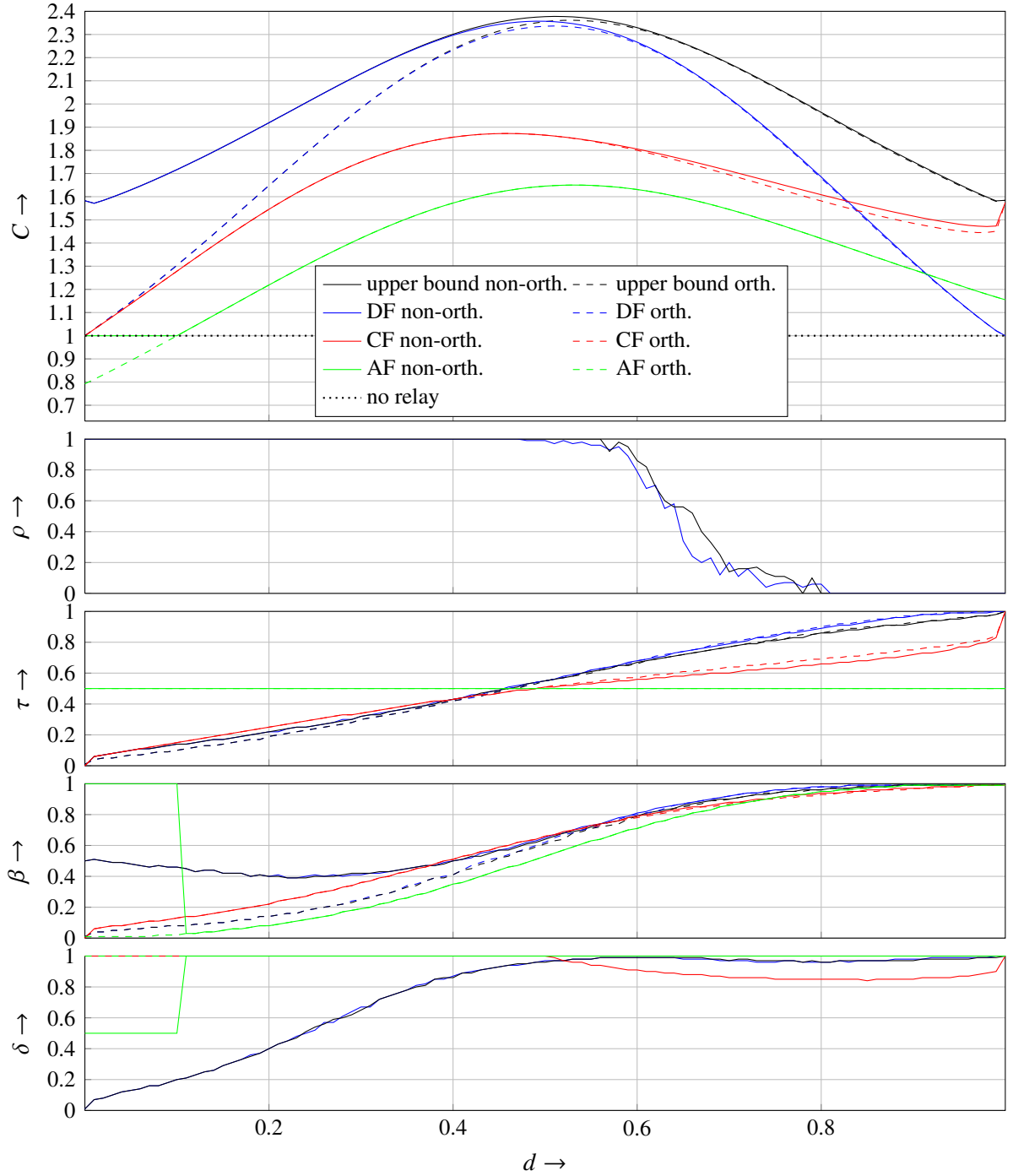


Figure 3.5: Achievable rates, upper capacity bound and optimal parameters ρ and τ versus source relay distance d with pathloss exponent $\alpha = 4$ and sum energy $E_{sum}^{non} = 1$.

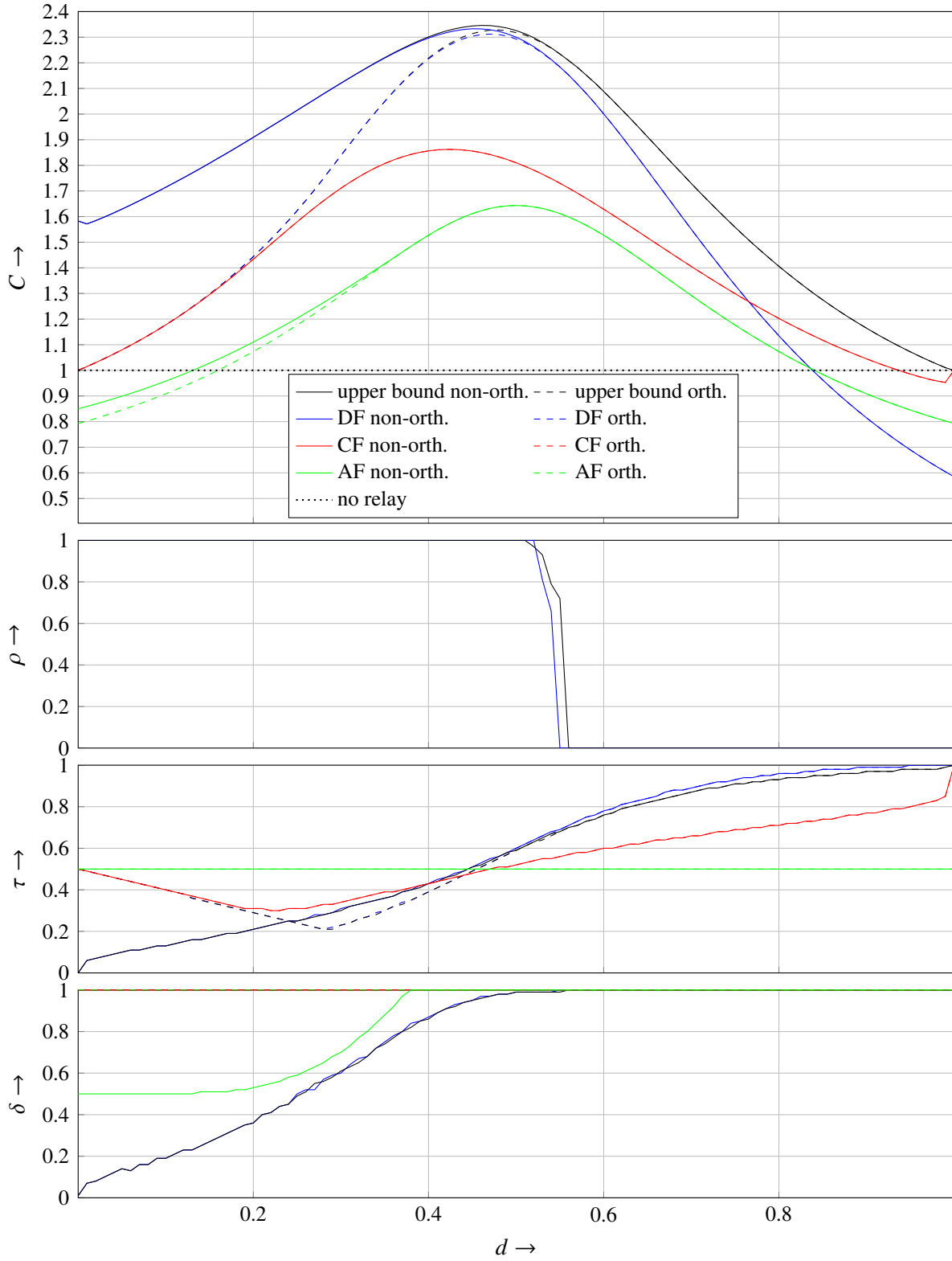


Figure 3.6: Achievable rates, upper capacity bound and optimal parameters ρ and τ versus source relay distance d with pathloss exponent $\alpha = 4$, sum energy $E_{sum}^{non} = 1$, and $\beta = 0.5$.

3.5.3 Individual Per Node Energy Constraint

By fixing β in the power allocation used in the previous section, the sum energy constraint becomes additionally an individual energy constraint. Fig. 3.6 shows results for $\beta = 0.5$, that is, source and relay have the same amount of energy available. The insights are very similar as in Fig. 3.5 but due to $\beta = 0.5$, the absolute values of the respective rates are lower in Fig. 3.6 for $d > 0.5$. For the relay being very close to the destination all relay protocols are worse than the direct transmission (except the upper bound). Comparing Fig. 3.5 and Fig. 3.6 in the range $d < 0.5$, DF and upper bound with non-orthogonal access behave very similar since $\beta \approx 0.5$ has turned out to be optimal. However, the behavior of orthogonal counterparts and CF changes such that the absolute values decrease. Hence, the gap between orthogonal and non-orthogonal DF increases while CF and orthogonal DF (and also the bound) perform very equal for $d < 0.2$. The beforehand small gain of non-orthogonal CF to orthogonal CF vanishes completely, i.e., non-orthogonal MAC does not pay off for an individual energy constraint per node. The same is true for DF when distributed beamforming is not reasonable. Finally, as $\beta = 1$ is not possible, non-orthogonal AF gets even worse than the direct link for $d < 0.1$.

3.6 Discussion

The 3-node relay channel, its capacity and achievable rates of specific relaying strategies (AF, DF, CF) have been studied from an information theoretic point of view for half-duplex relays whereby non-orthogonal and orthogonal MAC have been distinguished. Results reveal that all strategies are usually beneficial compared to a direct P2P transmission when the relay is placed between source and destination whereby AF is always outperformed by DF and CF. DF achieves the capacity bound when source and relay are at the same position while CF achieves the bound when relay and destination have the same position. For all other cases, the upper bound is not achieved. However, DF performs very close to this bound when the relay is placed close to the source. In such a scenario, the source applies distributed beamforming which achieves a high array gain. As beamforming requires phase knowledge at the source which cannot be guaranteed in practice, phase errors can degrade the achievable rate and, thus, lower the gain of non-orthogonal versus orthogonal MAC (cf. Ch. 5). For other relay positions, beamforming is no longer optimal and the source will either do a mixture of beamforming and superposition coding or only superposition coding (depending on ρ) whereby the latter can also be applied for CF which is usually only superior when the relay is very close to the destination. As the source's strategy for DF is the superposition of two independently encoded messages, one could think of a multilevel code with two levels where ρ denotes the power allocation between these levels. Therefore, the next chapter generalizes this perspective into a superposition multilevel code

which combines this classic strategy with superposition modulation to achieve an additional shaping gain as well as a low complex joint destination demapper. Finally, it turned out that non-orthogonal channel access (excluding array gain due to beamforming) does only pay off for a per node power constraint but not with respect to transmit energy.

Chapter 4

Superposition Multilevel Coding for Decode and Forward

From the previous chapter it is known that, for DF in the MAC phase, the source has the degree of freedom to choose the amount ρ^2 of power P_{S2} used for distributed beamforming whereby the residual power $(1 - \rho^2)P_{S2}$ is used to superpose an additional signal with independent information (superposition coding). Hence, ρ determines the correlation of source's and relay's signal in the second time slot. This superposition may be alternatively interpreted as two levels of a multilevel code where ρ determines the power allocation between the two layers of that code. Due to the inherent superposition of the source and the relay signal at the destination, it is natural to use superposition modulation to design such a multilevel code (instead of set partitioning as in Sec. 2.3.3). The generalization to more than two levels allows to make use of the inherent shaping gain of superposition modulation as explained in Sec. 2.3.4. During the work on this thesis, this concept has been published in [KK15b] as superposition multilevel coding and will be explained in the following: Sec. 4.1 will sketch the idea how a general superposition multilevel code arises from the original information theoretic concept. In Sec. 4.2, the general superposition multilevel code will be modified such that both messages can be jointly decoded at the destination.¹ Finally, Sec. 4.3 will show results with respect to the Max-Flow Min-Cut (MFMC) theorem.

4.1 General Distributed Superposition Multilevel Code

Considering the MAC phase of DF, the relay has already decoded the source's message and will transmit the re-encoded signal x_R (cf. Sec. 3.3.2), the source will transmit a superposition of x_R and new information \tilde{x}_{S2} , and the destination receives a superposition of all those signals

¹With complexity comparable to a single decoder while avoiding successive interference cancellation.

as described in (3.12a). Graphically this procedure is illustrated in Fig. 4.1 where source and relay processing are separated by the blue shaded boxes. The red shaded area can be interpreted as a joint mapper of a multilevel code with two levels where each of the levels belongs to either messages u_1 or u_2 .

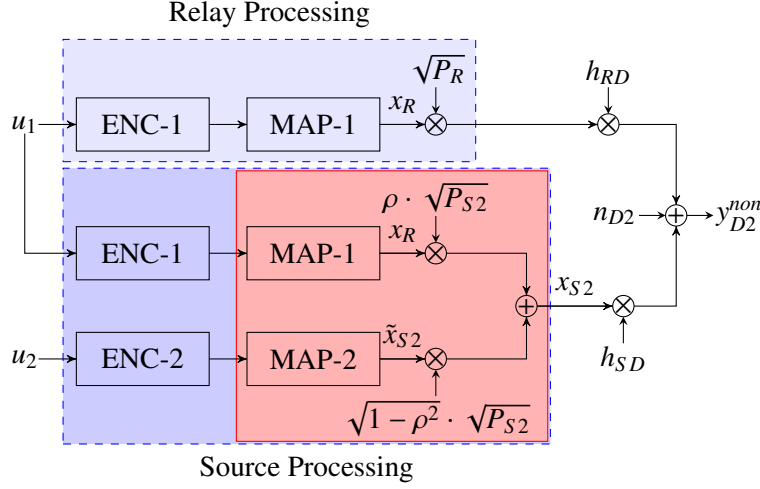


Figure 4.1: Conventional processing at source and relay for the MAC phase. Due to error free decoding, the relay has complete knowledge about u_1 .

Using superposition modulation, both messages may be multiplexed into more levels to create any desirable symmetric mapping for x_R , x_{S2} or the joint mapping of both after physical superposition at the destination. Doing so, transforms the structure in Fig. 4.1 into a superpo-

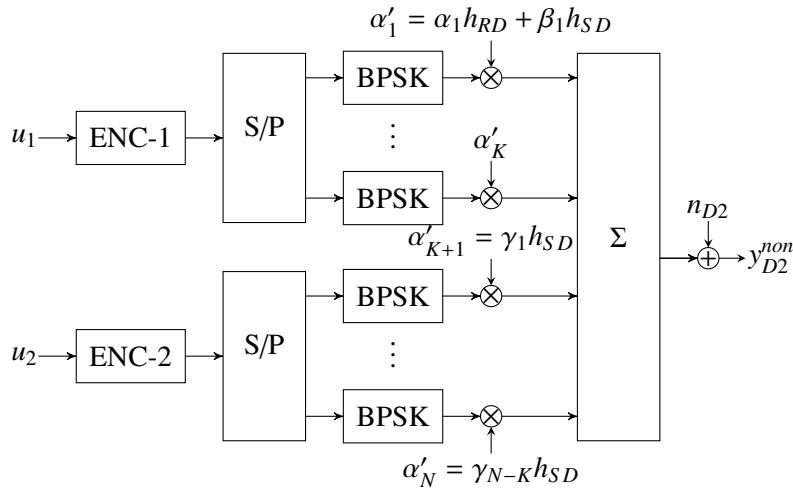


Figure 4.2: Concept of Distributed Superposition Multilevel Coding for the classical relay channel.

sition multilevel code (cf. Fig. 4.2) whose N levels have virtual amplitudes α'_n whereby the first K layers belong to message u_1 and the residual $N - K$ to u_2 . These virtual amplitudes α'_n have to be understood from the perspective of the superposition at the destination, that is, α'_n depends

on physically allocated level amplitudes α_n, β_n , and γ_n as well as the path loss coefficients h_{SR} and h_{SD} which have to be known at the source anyway to apply distributed beamforming.

$$\alpha'_n = \alpha_n h_{RD} + \beta_n h_{SD}, \quad 1 \leq n \leq K, \quad (4.1a)$$

$$\alpha'_n = \gamma_{n-K} h_{SD}, \quad K+1 \leq n \leq N, \quad (4.1b)$$

The physical amplitudes have to be allocated according to a specific power allocation scheme as described in Sec. 2.3.4 whereby

- α_n belongs to u_1 at the relay restricted by P_R ,
- β_n belongs to u_1 at the source restricted by $\rho^2 P_{S2}$, and
- γ_n belongs to u_2 at the source restricted by $(1 - \rho^2) P_{S2}$.

$$\sum_{n=1}^K \alpha_n^2 = P_R \quad \sum_{n=1}^K \beta_n^2 = \rho^2 P_{S2} \quad \sum_{n=K+1}^N \gamma_{n-K}^2 = (1 - \rho^2) P_{S2} \quad (4.2)$$

Depending on the power allocation scheme, α_n, β_n and γ_n are used to create the signals x_R and \tilde{x}_{S2} at source and relay. Thereby, the upper branch in Fig. 4.2 represents the distributed beamforming to transmit u_1 via x_R from S and R jointly to D while the lower branch illustrates the direct transmission of u_2 via \tilde{x}_{S2} from R to D . For the results in Sec. 4.3, grouped power allocation is assumed, that is, N layers are divided into L groups of size G whereby K must be chosen to be a multiple of G . Then $L_1 = \frac{K}{G}$ and $L_2 = \frac{N-K}{G}$ denote the number of groups belonging to u_1 and u_2 , respectively.² According to (2.10), α_n, β_n , and γ_n are chosen to meet the power restrictions in (4.2). For this purpose, it is usually not necessary to calculate explicitly initial values α_1, β_1 , and γ_1 . Instead, one could start with any value in (2.10) and normalize all α_n according to (4.2) afterwards. For a better understanding of Sec. 4.2 however, these initial values are given in the following by plugging the exponential law of (2.10) in each of the power constraints of (4.2).

$$\alpha_1 = \sqrt{P_R} \left(\frac{4}{3} G \left(1 - \frac{1}{4^{L_1}} \right) \right)^{-1/2} \quad (4.3a)$$

$$\beta_1 = \rho \sqrt{P_{S2}} \left(\frac{4}{3} G \left(1 - \frac{1}{4^{L_1}} \right) \right)^{-1/2} \quad (4.3b)$$

$$\gamma_1 = \sqrt{(1 - \rho^2) P_{S2}} \left(\frac{4}{3} G \left(1 - \frac{1}{4^{L_2}} \right) \right)^{-1/2} \quad (4.3c)$$

Design parameters G, L_1 and L_2 have to be chosen appropriately as stated in Sec. 4.3 while ρ is chosen in accordance with Fig. 3.3. Given that the mappers in Fig. 4.1 are QAM-mappers,

²Obviously, different group sizes G_1 and G_2 are possible but not meaningful (cf. Sec. 2.3.4).

the superposition multilevel code in Fig. 4.2 can be considered as generalization since QAM-mappings can be created with superposition modulation as shown in Sec. 2.3.4. Furthermore, the additional degrees of freedom of the new scheme in Fig. 4.2 provide several advantages: Firstly, the inherent shaping gain allows to approach the theoretical DF rate for the AWGN channel presented in Sec. 3.5. Secondly, each level may be encoded with different codes which can improve the convergence behavior of practical schemes like BICM [TP10]. Finally, the demapping complexity at the destination can be reduced to that of a single data stream if power allocation can be done jointly given additional side constraints as introduced in the following section. In particular, u_1 and u_2 will be detected in parallel without applying successive interference cancellation.

4.2 Distributed Superposition Multilevel Code with Parallel Detection

In contrary to the previous section, where α_n , β_n , and γ_n are determined independently, parallel detection with low complexity requires that power allocation is done jointly for α_n , β_n , and γ_n . The aim is to create virtual amplitude levels α' fulfilling a specific power allocation which is assumed to be grouped power allocation as in (2.10) for the rest of the chapter. Therefore, α'_n must fulfill both (2.10) and (4.1) which creates an additional side condition to the power allocation of the physical amplitudes α_n , β_n , and γ_n . To visualize this condition, let us consider the transition from the last layer belonging to u_1 to the first layer of u_2 . According to (4.1b), the $K + 1$ th layer is

$$\alpha'_{K+1} = \gamma_1 h_{SD}. \quad (4.4a)$$

Assuming for the sake of simplicity a group size of $G = 1$, (2.10) delivers

$$\alpha'_{K+1} = \alpha'_1 \cdot 2^{-K} \quad (4.4b)$$

whereby $\alpha'_1 = \alpha_1 h_{RD} + \beta_1 h_{SD}$ can be plugged in according to (4.1a). Subsequent equating of (4.4a) and (4.4b) yields

$$\gamma_1 = \left(\alpha_1 \frac{h_{RD}}{h_{SD}} + \beta_1 \right) \cdot 2^{-K} \quad (4.4c)$$

which may be contrasting to the condition in (4.2). Hence, there is no guaranteed solution for all combinations of K , N , and ρ given a specific setup according to h_{SR} and h_{SD} . Anyway, if (4.4c) can be fulfilled together with (4.2), u_1 and u_2 can be detected in parallel as if they are one

multiplexed data stream.³ To derive an adapted joint power allocation, a couple of new variables need to be introduced. So far, the number of all layers N is divided into L groups of size G , whereby K layers are dedicated to u_1 and $N - K$ to u_2 . In contrary to the previous section, it is not necessary that K is a multiple of G which means that one group may be split, that is, a couple of layers may belong to u_1 and the remaining to u_2 . Hence,

- $L_1 = \left\lfloor \frac{K}{G} \right\rfloor$ is the number of full groups related to u_1 ,
- $G_1 = K \bmod G$ is the size of the split group related to u_1 ,
- $G_2 = G - G_1$ is the size of the split group related to u_2 ,
- $L_2 = L - L_1 - 1$ is the number of full groups related to u_2 .⁴

According to these definitions, (4.2) can be rewritten as follows.

$$G \sum_{l=1}^{L_1} \alpha_l^2 + G_1 \alpha_{L_1+1}^2 = P_R \quad (4.5a)$$

$$G \sum_{l=1}^{L_1} \beta_l^2 + G_1 \beta_{L_1+1}^2 = \rho^2 P_{S2} \quad (4.5b)$$

$$G_2 \gamma_1^2 + G \sum_{l=1}^{L_2} \gamma_{l+1}^2 = (1 - \rho^2) P_{S2} \quad (4.5c)$$

As before, plugging the exponential law of (2.10) in (4.5), the following system of equations arises for the initial physical amplitudes of the power allocation.

$$\alpha_1 = \sqrt{P_R} \left(\frac{4}{3} G \left(1 - \frac{1}{4^{L_1}} \right) + G_1 4^{-L_1} \right)^{-1/2} \quad (4.6a)$$

$$\beta_1 = \rho \sqrt{P_{S2}} \left(\frac{4}{3} G \left(1 - \frac{1}{4^{L_1}} \right) + G_1 4^{-L_1} \right)^{-1/2} \quad (4.6b)$$

$$\gamma_1 = \sqrt{(1 - \rho^2) P_{S2}} \left(G_2 + \frac{G}{3} \left(1 - \frac{1}{4^{L_2}} \right) \right)^{-1/2} \quad (4.6c)$$

Generalizing (4.4c) to greater group sizes, adds a fourth equation.

$$\gamma_1 = \left(\alpha_1 \frac{h_{RD}}{h_{SD}} + \beta_1 \right) \cdot 2^{-L'} \quad (4.7)$$

Due to the additional condition (4.7) on γ_1 , (4.6) cannot be independently solved as (4.3) in the previous section. Instead, (4.7) and (4.6) form a system of equations with four equations and

³Obviously, every superposition can be detected in parallel without successive interference cancellation. The demapping complexity for superposition modulation however, is much lower than that of conventional schemes like the superposition of two QAMs with comparable performance [Wo11].

⁴The subtraction of 1 is dedicated to the split group. For the case that K is a multiple of G , there will still be one “split” group related to u_2 .

four unknowns $\alpha_1, \beta_1, \gamma_1$, and ρ . Actually (if ρ is chosen as before), this system of equations is over-determined due to (4.7) such that the solution space is narrowed tremendously. Especially due to the dependence on path loss coefficients h_{SR} and h_{SD} , there is almost no solution of (4.6) for a specific ρ . As a slight (or even a huge, cf. Fig. 4.3) deviation from the optimal ρ however might cause only a small degradation of the final transmission rate, $\rho \in [0, 1]$ is declared as unknown to obtain a meaningful solution space for a couple of relevant parameter choices. Subject to Sec. 2.3.4, meaningful parameter choices are $G \in \{4, 6\}$ (group size doubled to create complex signal space) and $L \leq 5$ with $K \leq G \cdot L$ in order that demapping complexity does not explode.⁵ Due to the restricted domain of $\rho \in [0, 1]$, there are still a lot of cases according to h_{SR} and h_{SD} without a solution. Nonetheless, especially for larger bit loads, it is possible to find reasonable values close to the theoretically optimal ρ as shown in Table 4.1.

Table 4.1: Values of ρ for different combinations of K and L for $G = 6$ (respectively 3 layers in I and Q), $d = 0.4$, and $P_{S1} = P_{S2} = P_R = P = 1$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 9$	0.176	-	-	-	-	-
$K = 10$	0.488	0.047	-	-	-	-
$K = 11$	0.755	0.377	0.275	0.249	0.242	0.241
$K = 12$	1.000	0.644	0.554	0.531	0.525	0.524
$K = 13$	-	0.706	0.617	0.595	0.589	0.588
$K = 14$	-	0.767	0.679	0.657	0.652	0.650
$K = 15$	-	0.827	0.740	0.718	0.713	0.711
$K = 16$	-	0.885	0.799	0.778	0.772	0.771
$K = 17$	-	0.943	0.858	0.836	0.831	0.830
$K = 18$	-	1.000	0.915	0.894	0.889	0.887
$K = 19$	-	-	0.929	0.908	0.903	0.902
$K = 20$	-	-	0.944	0.923	0.917	0.916
$K = 21$	-	-	0.958	0.937	0.932	0.930
$K = 22$	-	-	0.972	0.951	0.946	0.944
$K = 23$	-	-	0.986	0.965	0.960	0.958
$K = 24$	-	-	1.000	0.979	0.974	0.972
$K = 25$	-	-	-	0.983	0.977	0.976
$K = 26$	-	-	-	0.986	0.981	0.979
$K = 27$	-	-	-	0.990	0.984	0.983
$K = 28$	-	-	-	0.993	0.988	0.986
$K = 29$	-	-	-	0.997	0.991	0.990
$K = 30$	-	-	-	1.000	0.995	0.993

⁵From Sec. 2.3.4 it is known that $G \in \{2, 3\}$ are sufficiently large to achieve an almost optimal shaping gain for a real signal space. For complex signal spaces this holds for each dimension so that group size is doubled whereby half of the layers belongs to the I and the other half to the Q component.

4.3 Achievable Rates for Superposition Multilevel Coding

The following results are obtained similar as in Sec. 3.5 by means of the MFMC theorem for the same line setup (S and D fixed at 0 and 1 with varying $R\forall d \in [0, 1]$) and transmit powers ($P_{S1} = P_{S2} = P_R = P$). Unfortunately, lacking a closed form solution, the cut-sets have to be solved numerically as explained in App. C.2. Despite the above presented superposition multilevel code, QAM inputs are considered for comparison. Modulation orders and the number of bit levels, i.e., the cardinality of the input alphabet, are chosen such that respective link capacities are not saturated. Anyway, due to complexity issues of the numerical evaluation as well as for practical relevance the maximum modulation order is limited to $m \leq 10$ for QAM and to $L_1, L_2 \leq 4$ and $L_1, L_2 \leq 5$ for superposition modulation with $G = 6$ and $G = 4$, respectively. This limitation will cause a loss whenever the capacity is beyond the input entropy (saturation) which is surely the case for relay positions very close to the source or to the destination. Hence, the range of d is restricted to $\{0.1, \dots, 0.9\}$.⁶

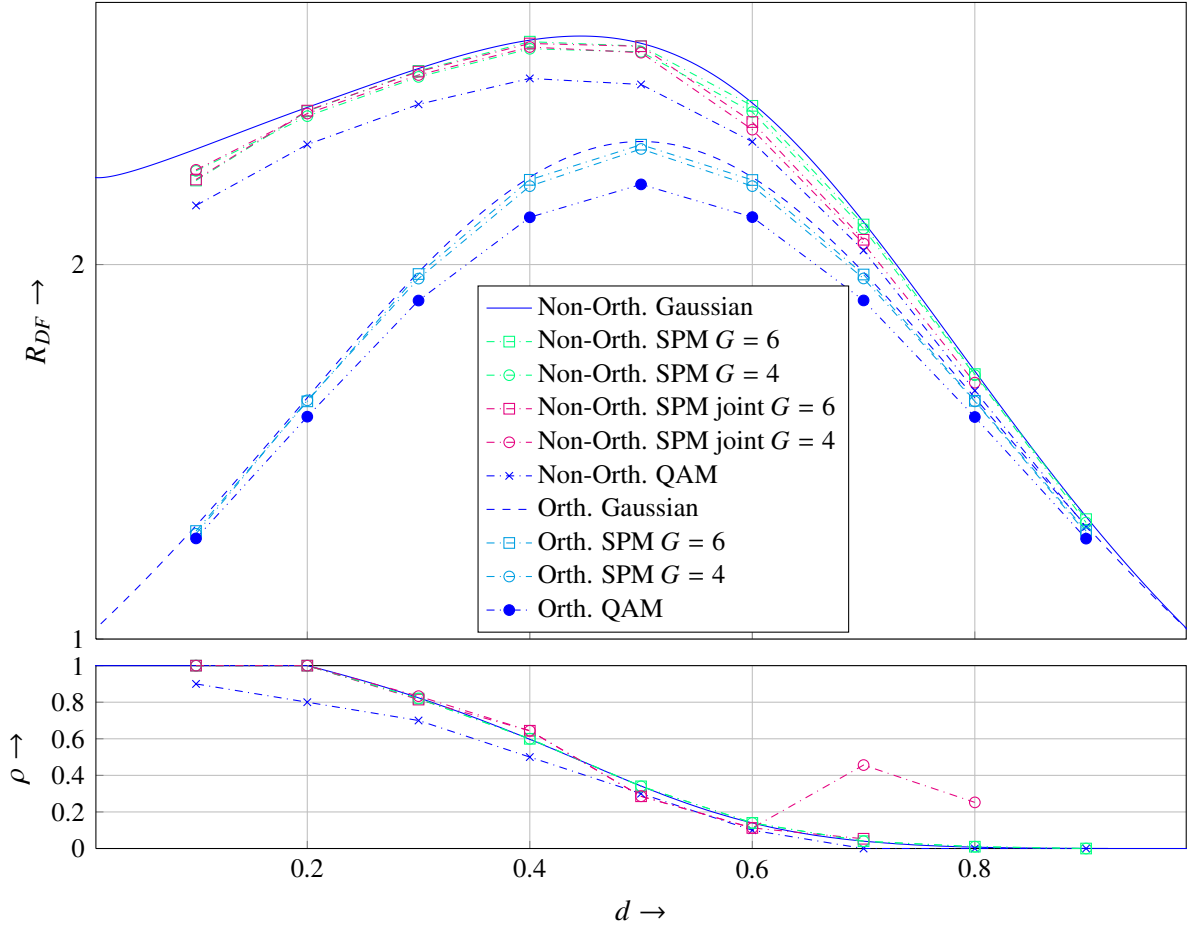


Figure 4.3: Achievable rates and optimal correlation coefficient ρ versus source relay distance d for DF for superposition modulation and QAM with $\alpha = 4$ and transmit power $P = 1$.

⁶The coarse resolution is owed to the high complexity of the numerical evaluation of the cut-sets.

Fig. 4.3 shows the achievable rates for DF versus relay position d for $P = 1$. Similar as in Fig. 3.3, solid and dashed lines belong to a Gaussian input. Additional plots correspond to discrete inputs, either state-of-the-art QAM or the above introduced superposition multilevel code. Firstly, the general superposition multilevel code closely reaches the achievable rate of the Gaussian input except for non-orthogonal channel access at $d = 0.1$ and clearly outperforms classical QAM in the total range. As expected, a group size of $G = 6$ leads to a slightly better performance than $G = 4$ except for $d = 0.1$. The distinct gap at $d = 0.1$ is caused by saturation (cf. Fig. 2.20) on the source-relay link due to restriction on L which limits input entropy: As $G = 4$ allows higher $L = 5$ compared to $L = 4$ for $G = 6$, it is beneficial at this point. Comparing the general superposition modulated input to the one with parallel detection for $G = 6$ reveals that the deviation from the optimal ρ causes only minor performance losses for $0.1 \leq d \leq 0.7$. Unfortunately, there are no results for $d > 0.7$ ($d > 0.8$) due to construction impossibility according to Table E.7 and Table E.8 which supports only very high values for ρ or an infeasibly high number $L > 4$ of differing power levels. A similar behavior is visible for $G = 4$ even for $d = 0.7$ and $d = 0.8$ where the deviation of ρ is quite huge.

4.4 Discussion

In this chapter, a superposition multilevel code, distributed between source and relay, has been designed which generalizes the classical information theoretic principle of superposition coding. This multilevel code provides a shaping gain which allows performance close to the theoretically achievable rate of DF (cf. Fig. 4.3). Following additionally a more restrictive design, the joint mapping of this superposition modulation, can be demapped in parallel with low complexity at the destination whereby the performance loss is only marginal. Both schemes clearly outperform classical QAM while having lower complexity as gray-labeled QAM especially for the advanced design. The results for QAM are reused in the next chapter which analyzes the influence of imperfect carrier phase synchronization on distributed beamforming.

Chapter 5

Decode and Forward for Perfect and Imperfect Carrier Synchronization

This chapter investigates the Decode-and-Forward (DF) relay protocol taking into account practical aspects like the influence of imperfect carrier synchronization on distributed beamforming (cf. Sec. 3.5). In particular, phase deviations from a perfectly constructive superposition are investigated while all other effects are assumed to be perfectly estimated and compensated at the receiver. Firstly, an appropriate phase error model will be introduced in terms of a random variable in Sec. 5.1 and Sec. 5.2. Secondly, the influence of this random variable will be investigated with respect to the achievable rates according to the MFMC theorem whereby results for QAM are given in Sec. 5.5.1. Finally, similar results are obtained for the BICM scheme presented in Sec. 2.3.5 whereby rate allocation for partial link transmissions is done based on Fig. 2.22. The derivation of the phase error and respective results have been published in [KK15c].

5.1 Distributed Beamforming with Phase Errors

Distributed beamforming is only meaningful for DF and in scenarios where R is closer to S than to D (cf. Fig. 3.3). For the extreme case, where R is very close to S , all available transmit power will be used for beamforming which is very prone to phase differences between S and R . Such phase differences occur due to two reasons: Firstly, the oscillators used at S and R exhibit production tolerances, and, thus cannot produce exactly the same carrier frequency. Secondly, due to different channels h_{RD} and h_{SD} , the source needs knowledge about the channel phases to adapt its transmit signal such that the superposition at the destination is constructive. One option would be an estimation at the destination with subsequent feedback to the source. As the estimation is erroneous in the presence of noise, phase errors occur and maybe even outdated due to the caused delay. The latter effect may be compensated by an extrapolation based on

previous observations. To analyze the influence of such phase errors, a random variable Φ with realization ϕ is introduced into (3.12a) as follows.

$$y_{D2}^{non} = h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + (h_{SD} \sqrt{P_{S2}} \cdot \rho \cdot e^{j\phi} + h_{RD} \sqrt{P_R}) \cdot x_R + n_{D2} \quad (5.1)$$

The phase rotation $e^{j\phi}$ models the erroneous phase difference between the source's and the relay's component x_R and leads to a non-coherent superposition at the destination D . Despite a rotation of the signal space, this phase rotation will decrease the SNR at D which in worst case corresponds to a complete cancellation of the signal.

5.2 Random Phase Error Model

A very common way for carrier phase synchronization is to use a Phase Locked Loop (PLL). Let the input signal with amplitude A , carrier frequency f_c , and phase θ to this PLL be [JBS00]

$$y_{PLL}(t) = \sqrt{A} \cos(2\pi f_c t + \theta) + n(t) \quad (5.2)$$

where $n(t)$ is the AWGN at time instance t with power spectral density N_0 [Vit63]. The loop is assumed to be in tune and both transmitter and receiver oscillator are assumed to be perfectly stable. Then, the pdf of the modulo 2π reduced phase error $\phi = \theta - \hat{\theta}$ with estimate $\hat{\theta}$ can be determined analytically for a first order loop. The derivation can be found in detail in [Vit63]: Viterbi solved the problem using continuous random walk or Fokker-Planck techniques. The final result for the pdf of phase error ϕ is [JBS00, p. 662]

$$p_\Phi(\phi) = \frac{\exp(\gamma \cos \phi)}{2\pi I_0(\gamma)} \quad |\phi| \leq \pi \quad (5.3)$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind and $\gamma = \frac{A^2}{N_0 B_L}$ the SNR of the PLL with loop bandwidth B_L . This loop SNR turns out to be the parameter which steers the impact of the phase error as illustrated in Fig. 5.1:

- For γ approaching zero, (5.3) becomes a uniform distribution.
- For γ being large, (5.3) becomes approximately Gaussian (dashed in Fig. 5.1) as shown in the following [Vit63].
- For γ approaching infinity, (5.3) yields a Dirac impulse at $\phi = 0$.

When the loop SNR γ is sufficiently high, $I_0(\gamma)$ is approximately [Vit63]

$$\frac{\exp(\gamma)}{\sqrt{2\pi\gamma}} \quad (5.4)$$

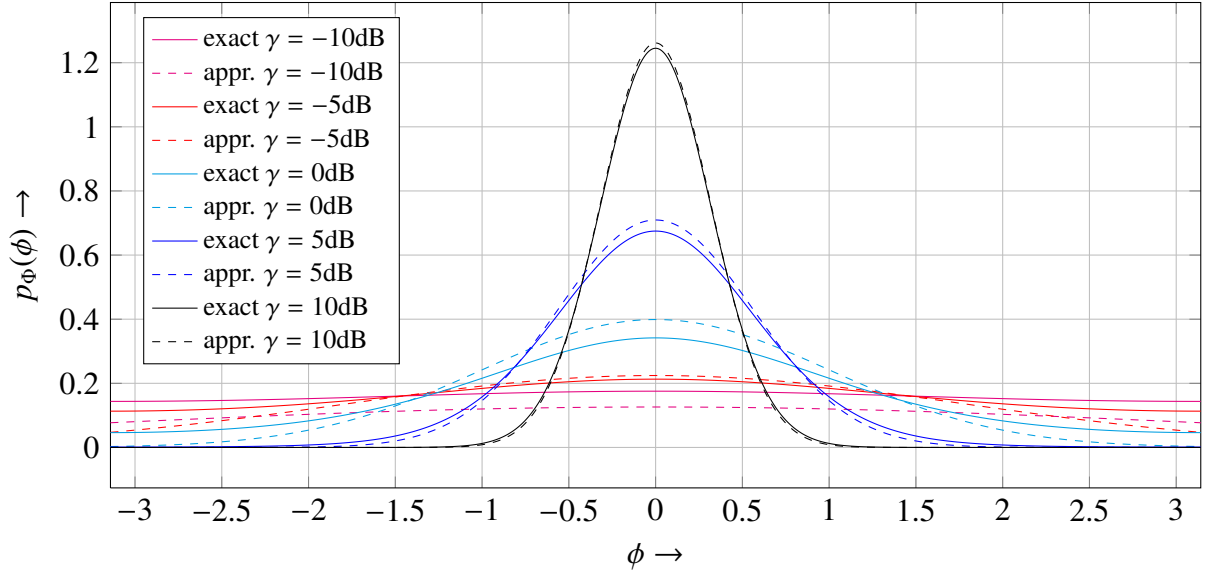


Figure 5.1: Exact (cf. (5.3)) and approximated (cf. (5.6)) $p_\Phi(\phi)$ for different loop SNRs γ .

which delivers

$$p_\Phi(\phi) = \left(\frac{2\pi}{\gamma}\right)^{-1/2} \exp(\gamma \cos \phi - 1) \quad |\phi| \leq \pi. \quad (5.5)$$

Expanding $\cos \phi$ into a Taylor series and cutting after the second order term since higher order terms have only little effect, delivers [Vit63, JBS00]

$$p_\Phi(\phi) \approx (2\pi\sigma_\Phi^2)^{-1/2} \cdot \exp(-\phi^2/2\sigma_\Phi^2) \quad (5.6)$$

which is indeed a Gaussian $\phi \sim \mathcal{N}(0, \sigma_\Phi^2)$ with $\sigma_\Phi^2 = \gamma^{-1}$.

5.3 Max Flow Min Cut Theorem with Phase Errors

According to (5.1) only terms $I(X_{S2}, X_R; Y_{D2}^{non})$ and $I(X_{S2}; Y_{D2}^{non}|X_R)$ in (3.13) are influenced by phase error ϕ . From (C.21) it becomes obvious that pdf $p_{Y_{D2}^{non}|X_R \tilde{x}_{S2}}(y_{D2}^{non}|x_R, \tilde{x}_{S2})$ is of interest. Therefore, (5.1) can be rewritten as

$$y_{D2}^{non} = h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + h_{RD} \sqrt{P_{RX_R}} + h_{SD} \sqrt{P_{S2}} \cdot \rho \cdot e^{j\phi} x_R + n_{D2} \quad (5.7a)$$

$$\stackrel{(a)}{=} \underbrace{h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + h_{RD} \sqrt{P_{RX_R}}}_{\tilde{x} = r_0 \cdot e^{j\varphi_0}} + e^{j\phi} \underbrace{\left(h_{SD} \sqrt{P_{S2}} \rho \cdot x_R + e^{-j\phi} \cdot n_{D2} \right)}_{\tilde{y}_{D2} = |\tilde{y}_{D2}| \cdot e^{j(\angle \tilde{y}_{D2})}} \quad (5.7b)$$

$$\stackrel{(b)}{=} \tilde{x} + |\tilde{y}_{D2}| \cdot e^{j(\phi + \angle \tilde{y}_{D2})} \quad (5.7c)$$

where steps (a) and (b) are meaningful for the derivation of $p_{Y_{D2}^{non}|X_R\tilde{x}_{S2}}(y_{D2}^{non}|x_R, \tilde{x}_{S2})$. In (a), $e^{j\phi}$ is placed outside the brackets to introduce the auxiliary variable $\tilde{y}_{D2} = h_{SD} \sqrt{P_{S2}} \rho \cdot x_R + e^{-j\phi} \cdot n_{D2}$ with pdf $p_{\tilde{y}_{D2}|X_R}(\tilde{y}_{D2}|x_R)$ which is a complex Gaussian $CN(h_{SD} \sqrt{P_{S2}} \cdot \rho \cdot x_R, 1)$ since the statistics of noise n_{D2} are rotationally invariant. Furthermore, $\tilde{x} = h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + h_{RD} \sqrt{P_R} x_R$ is substituted for the sake of clarity since it depends only on non-random or given variables. Step (b) uses polar coordinates to highlight that $e^{j\phi}$ only changes the phase $\varphi_{\tilde{y}_{D2}} = \angle \tilde{y}_{D2}$ of \tilde{y}_{D2} but not its magnitude $r_{\tilde{y}_{D2}} = |\tilde{y}_{D2}|$. Most importantly, the sum of angles corresponds to a convolution of the corresponding phases of the pdfs $p_{\Phi}(\phi) *_{\varphi} p_{\tilde{y}_{D2}|X_R}(r_{\tilde{y}_{D2}}, \varphi_{\tilde{y}_{D2}}|x_R)$. The summation with $\tilde{x} = r_0 \cdot e^{j\varphi_0}$ corresponds to a convolution with a Dirac impulse which is

$$\frac{\delta(r - r_0, \varphi - \varphi_0)}{r} \quad \text{with} \quad \begin{aligned} r_0 &= \left| h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + h_{RD} \sqrt{P_R} x_R \right| \\ \varphi_0 &= \angle \left(h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + h_{RD} \sqrt{P_R} x_R \right) \end{aligned} \quad (5.8)$$

in polar coordinates. The final pdf in polar coordinates ($r = |y_{D2}|$, $\varphi = \angle y_{D2}$) is then

$$p_{Y_{D2}^{non}|X_R\tilde{x}_{S2}}(r, \varphi|x_R, \tilde{x}_{S2}) = \frac{\delta(r - r_0, \varphi - \varphi_0)}{r} * p_{\Phi}(\phi) *_{\varphi} p_{\tilde{y}_{D2}|X_R}(r, \varphi|x_R), \quad (5.9)$$

which can be solved numerically by means of the FFT, that is, a convolution of pdfs becomes the product of the corresponding characteristic functions. Let us firstly assume that there are no phase errors, that is, loop SNR $\gamma = \infty$ dB, i.e., $p_{\Phi}(\phi)$ becomes a Dirac impulse at zero which has no influence in (5.9). Then, $p_{Y_{D2}^{non}|X_R\tilde{x}_{S2}}(r, \varphi|x_R, \tilde{x}_{S2})$ is simply a complex Gaussian with $CN(h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + (h_{SD} \sqrt{P_{S2}} \cdot \rho + h_{RD} \sqrt{P_R}) \cdot x_R, 1)$ at the position of the superposed transmit signals. For $\phi \neq 0$ however, $e^{j\phi}$ causes a smearing of the usually bell shaped Gaussian in the phase dimension (due to convolution with $p_{\Phi}(\phi)$, cf. Fig. 5.1) as shown in Fig. 5.2 where the constant transition caused by the Dirac impulse $\frac{\delta(r - r_0, \varphi - \varphi_0)}{r}$ is omitted ($r_0 = 0$, $\varphi_0 = 0$) for the sake of clarity.¹ For a high loop SNR $\gamma = 5$ dB (cf. Fig. 5.2(a)), that is, low variance σ_{Φ}^2 , the effect is quite low. Furthermore, both the exact (cf. (5.3)) and the approximated (cf. (5.6)) pdf $p_{\Phi}(\phi)$ deliver a very similar result. For moderate and low loop SNRs of $\gamma = 0$ dB (cf. Fig. 5.2(b)) and $\gamma = -5$ dB (cf. Fig. 5.2(c)) however, the effect of this phase smearing, caused by the convolution in the phase dimension, is much stronger. Especially in Fig. 5.2(c) the distribution equals more and more a uniform circle because $p_{\Phi}(\phi)$ develops into a uniform distribution for $\gamma \rightarrow \infty$. Comparing again exact and approximated $p_{\Phi}(\phi)$ in Fig. 5.2(b) and Fig. 5.2(c), a reasonable difference becomes visible whereby the approximation is too optimistic.

¹In Fig. 5.2, there is rotation around zero (with radius $h_{SD} \sqrt{P_{S2}} \rho \cdot |x_R|$) which would be usually around $\tilde{x} = h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + h_{RD} \sqrt{P_R} x_R$ (cf. (5.7b)).

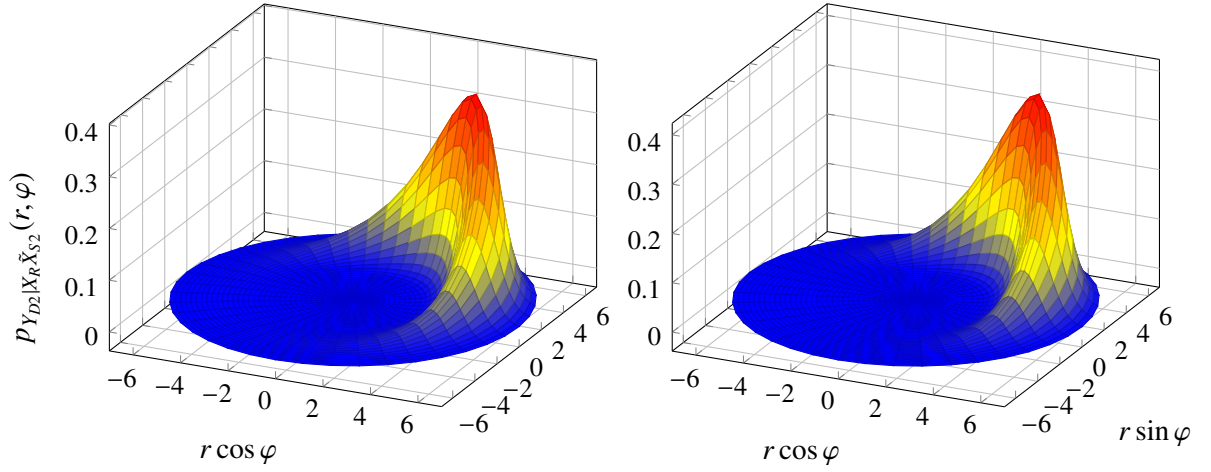
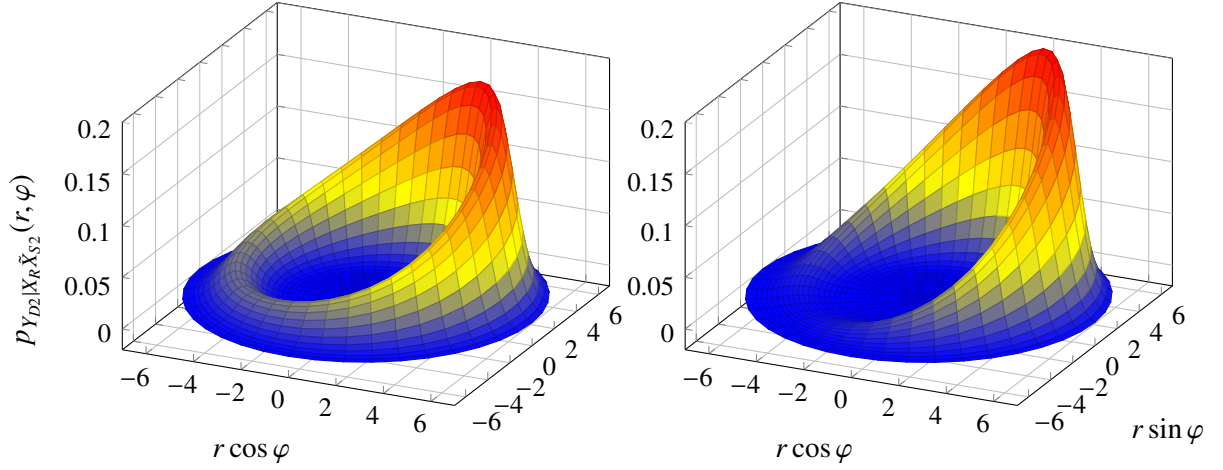
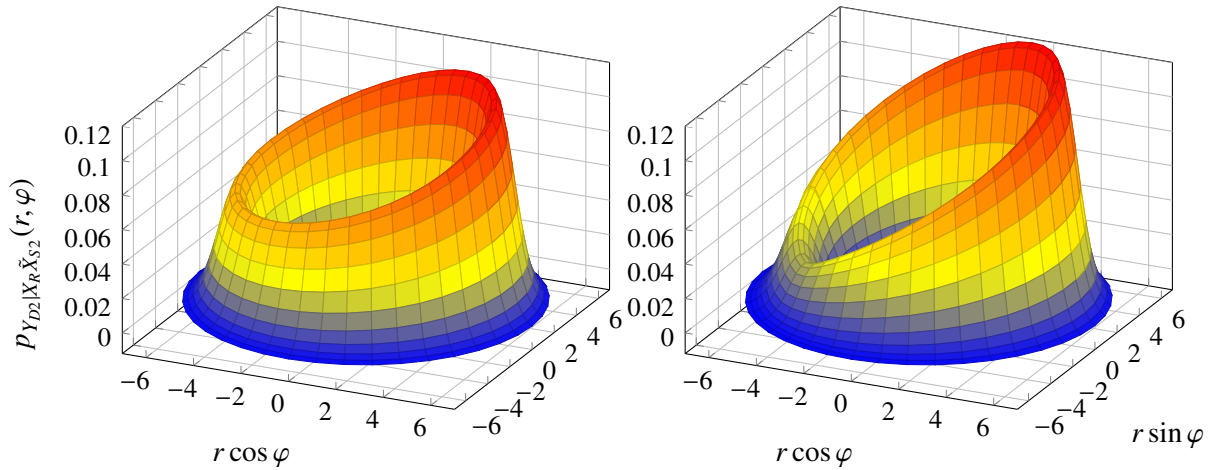
(a) Exact (left) and approximated (right) distribution for loop SNRs $\gamma = 5$ dB.(b) Exact (left) and approximated (right) distribution for loop SNRs $\gamma = 0$ dB.(c) Exact (left) and approximated (right) distribution for loop SNRs $\gamma = -5$ dB.

Figure 5.2: $p_{Y_{D2}|X_R\tilde{X}_{S2}}(r, \varphi)$ with $h_{SD} \sqrt{P_{S2}\rho} \cdot |x_R| = 5 \cdot e^{j\frac{\pi}{4}}$, $r_0 = 0$, $\varphi_0 = 0$ (cf. (5.9)) for exact (cf. (5.3)) and approximated (cf. (5.6)) $p_\Phi(\phi)$ and different loop SNRs γ .

5.4 Rate Allocation for Practical System with Phase Errors

In Ch. 2, Bit Interleaved Coded Modulation (BICM) is introduced with 40 Modulation and Coding Schemes (MCSs) for Point-to-Point (P2P) transmission whereby, for each MCS, the achievable rate R_b and the corresponding threshold SNRs have been determined via Monte Carlo simulation for the AWGN channel. Here, these results are used to assign such a rate R_b to each partial transmission in the 3-node relay channel for DF. Briefly summarizing the partial transmissions, there is a broadcast of x_{S1} according to (3.10), a forwarding of x_R from R to D according to (3.12b) or (3.12a) depending on the channel access being orthogonal or non-orthogonal, and maybe a virtual transmission of \tilde{x}_{S2} for non-orthogonal MAC with $\rho < 1$. The partial transmission rates $R_b^{x_{S1}}$, $R_b^{x_R}$, and $R_b^{\tilde{x}_{S2}}$ can then be determined straightforward with the corresponding SNRs and the help of the results from Sec. 2.4. Finally, the totally achievable rate R_{DF} for DF and also τ can be calculated from this partial rates. For the sake of comparison, orthogonal channel access will be considered as well since it does not require phase knowledge at all.

5.4.1 Orthogonal Channel Access

The overall goal is to encode x_{S1} with a rate $R_b^{x_{S1}}$ of a specific MCS such that the destination can decode successfully. Usually, the destination decodes y_{D1} and y_{D2}^{orth} jointly to detect x_{S1} but due to the restriction of DF, x_{S1} must be encoded such that the relay can successfully decode the original message from y_R . This is indeed a P2P transmission from S to R with

$$\text{SNR} = |h_{SR}|^2 \cdot P_{S1} \quad (5.10)$$

at the relay according to (3.10a). Given this SNR, the rate $R_b^{x_{S1}}$ can be found from Fig. 2.22. For the MAC phase where x_R is encoded with rate $R_b^{x_R}_{(orth)}$, the situation is different since the detection of the message u_1 encoded in x_{S1} and x_R does not only rely on y_{D2}^{orth} but also on y_{D1} which has been received in the broadcast phase. Due to this additional information, $R_b^{x_R}_{(orth)}$ can be higher than a comparable rate of a P2P transmission. Unfortunately, the determination of such a rate is quite tedious because, on one hand, this would require a concatenated turbo decoder with 4 component decoders which jointly decode y_{D1} and y_{D2}^{orth} at the destination. On the other hand, all 1600 combinations of the available 40 MCSs used to encode x_{S1} and x_R would have to be considered. Therefore, y_{D1} is omitted so that the rate $R_b^{x_R}_{(orth)}$ for transmission of x_R can be determined with the help of the link level simulations of Sec. 2.4 using the following SNR according to (3.12b).

$$\text{SNR} = |h_{RD}|^2 \cdot P_R \quad (5.11)$$

Despite the high complexity, some exemplary simulations considering the whole transmission chain with a joint decoder at D have been conducted to check how big the loss of neglecting y_{D1} really is: Especially for scenarios where the relay is close to the source, the decoding gain that y_{D1} offers in the joint decoder is really negligible. This is reasonable because in that case x_{S1} can be transmitted with high rate in short time due to the short distance of S and R . Hence, there is only a tiny information flow to D due to the much longer distance between S and D . In other scenarios however, e.g., when the relay is roughly in the middle between S and D , y_{D1} can provide a coding gain of about 2 dB.

After determining $R_b^{x_{S1}}$ and $R_b^{x_R}_{(orth)}$ via Fig. 2.22, the final rate for DF with orthogonal channel access is given by proper weighting of the just determined rates with the corresponding time slot length

$$R_{DF}^{orth} = \tau_{orth} \cdot R_b^{x_{S1}} = (1 - \tau_{orth}) \cdot R_b^{x_R}_{(orth)} \quad (5.12)$$

where τ_{orth} has to be chosen such that the information transmitted in first and second time slot is equal.

$$\tau_{orth} = \frac{1}{1 + \frac{R_b^{x_{S1}}}{R_b^{x_R}_{(orth)}}}, \quad (5.13)$$

This is directly indicated by (3.14) substituting the information terms by $R_b^{x_{S1}}$ and $R_b^{x_R}_{(orth)}$ as well as neglecting Y_{D1} .

5.4.2 Non-Orthogonal Channel Access

The transmission of x_{S1} in the first time slot does not differ from the orthogonal scheme and, thus, $R_b^{x_{S1}}$ can be determined as described above.² Furthermore, as beamforming can be applied due to the additional transmission of the source S in the second time slot, phase errors have to be taken into account. In a practical system as considered here, channel estimation is usually mandatory before signal detection. Hence, a signal rotation caused by constant phase errors will be compensated at the receiver. However, as the receiver observes only the superposition of two signals which cannot be separated, a phase difference between these two links cannot be corrected. This phase difference is then equivalent to a degraded SNR which is indeed an

²The time slot length of the broadcast phase will change for sure. The rate however does only depend on the SNR of the link from S to R .

SINR for $\rho < 1$. By inspection of (5.1), the SINR to determine the rate $R_{b(non)}^{x_R}$ to transmit x_R for non-orthogonal MAC is

$$\begin{aligned} \text{SINR} &= \frac{E\{|h_{SD} \sqrt{P_{S2}} \rho \cdot e^{j\phi} + h_{RD} \sqrt{P_R}\|^2\}}{h_{RD}^2(1 - \rho^2)P_{S2} + 1} \\ &= \frac{\rho^2 h_{SD}^2 P_{S2} + h_{RD}^2 P_R + 2\rho \cdot h_{SD} h_{RD} \sqrt{P_{S2} P_R} E\{\cos \phi\}}{h_{RD}^2(1 - \rho^2)P_{S2} + 1} \end{aligned} \quad (5.14)$$

with

$$\begin{aligned} E\{\cos \phi\} &= \int_{-\infty}^{\infty} \frac{\cos \phi}{2\pi\sigma_{\Phi}^2} \cdot \exp(-\phi^2/2\sigma_{\Phi}^2) d\phi \\ &= \exp(-\sigma_{\Phi}^2/2) \end{aligned} \quad (5.15)$$

which follows from $E\{g(x)\} = \int_{-\infty}^{\infty} g(x)p(x)dx$ [BSM05] with approximated pdf of Φ in (5.6). Please note that the intuitive way of taking the expectation of the totally achievable rate is not applied for the same reasons of neglecting the decoding gain of y_{D1} . For the cases with $\rho < 1$, successive interference cancellation has to be applied by the destination which means that the message in \tilde{x}_{S2} will be detected after subtracting the influence of x_R . Hence, the transmission of \tilde{x}_{S2} from S to D is virtually a simple P2P transmission and the transmission rate $R_b^{\tilde{x}_{S2}}$ can be determined with

$$\text{SNR} = h_{SD}^2(1 - \rho^2) \cdot P_{S2}. \quad (5.16)$$

Similar as in (5.12), the total rate can be determined by weighting of the just determined individual rates, whereby τ_{non} is chosen such that $\tau_{non} \cdot R_b^{x_{S1}} = (1 - \tau_{non}) \cdot R_{b(non)}^{x_R}$.

$$\tau_{non} = \frac{1}{1 + \frac{R_b^{x_{S1}}}{R_{b(non)}^{x_R}}} \quad (5.17)$$

In addition, \tilde{x}_{S2} is transmitted independently with $R_b^{\tilde{x}_{S2}}$ in the second time slot of length $(1 - \tau_{non})$. Hence, $(1 - \tau_{non}) \cdot R_b^{\tilde{x}_{S2}}$ has to be added either to $\tau_{non} \cdot R_b^{x_{S1}}$ or to $(1 - \tau_{non}) \cdot R_{b(non)}^{x_R}$.

$$R_{DF}^{non} = (1 - \tau_{non}) \cdot (R_{b(non)}^{x_R} + R_b^{\tilde{x}_{S2}}) \quad (5.18)$$

5.5 Achievable Rates for Decode and Forward with Phase Errors

As in Ch. 3, results are obtained for a line setup with $d_S = 0$, $d_R = d \in [0, 1]$ and $d_D = 1$. Transmit powers are set to $P_{S1} = P_{S2} = P_R = P$ according to the individual power constraint per node. Firstly, the results of the MFMC theorem for QAM inputs (cf. Sec. 4.3 and App. C.2) are extended by the influence of the introduced phase error. Secondly, Monte Carlo simulations have been conducted to obtain results for the BICM scheme presented in Sec. 2.3.5.

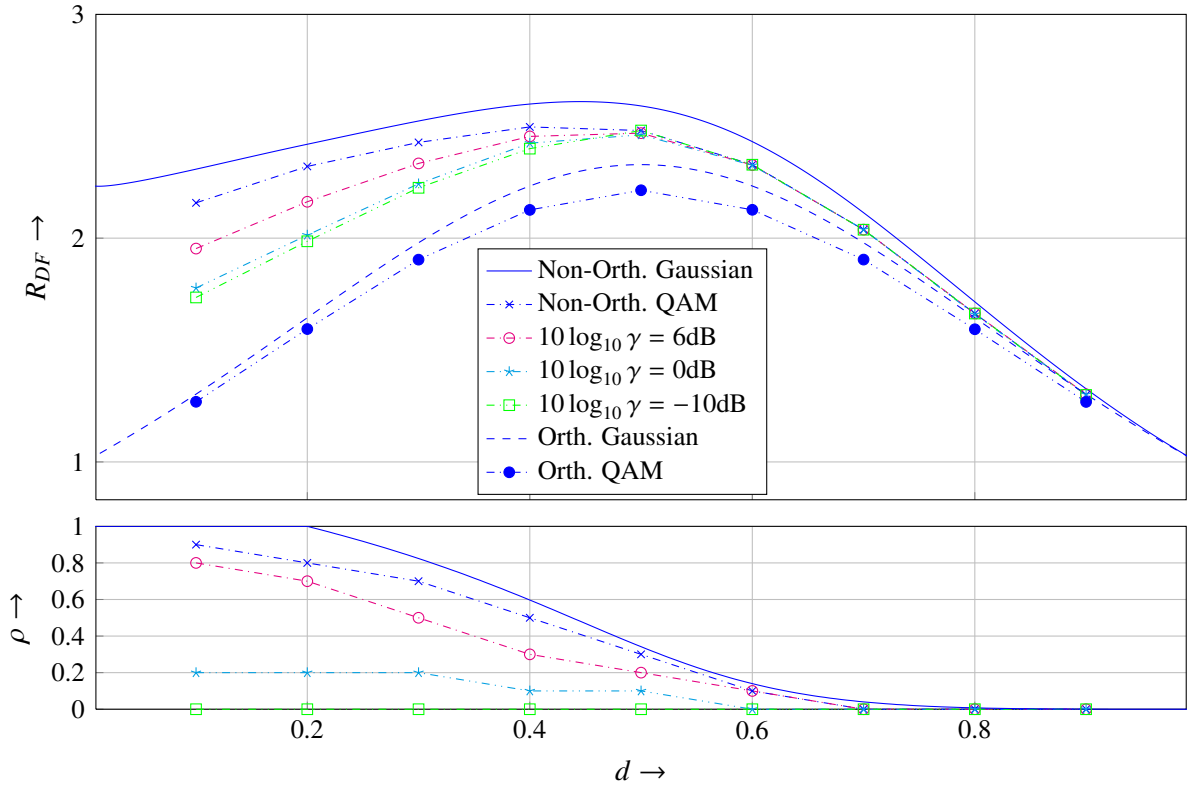


Figure 5.3: Achievable rates and optimal correlation coefficient ρ versus source relay distance d for DF with QAM input in the presence of phase errors with pathloss exponent $\alpha = 4$ and transmit power $P = 1$.

5.5.1 Results for Max Flow Min Cut Theorem

Fig. 5.3 shows the achievable rates R_{DF} vs. d for $P = 1$. Blue lines without markers correspond to Gaussian input as in Fig. 3.3 and blue lines with markers to QAM input as in Fig. 4.3. Additionally depicted rates suffer by the presence of phase errors according to the above specified model depending on loop SNR γ (cf. Fig. 5.2). As (5.9) has to be evaluated numerically in any case, the exact pdf (5.3) is used to model phase errors (cf. Fig. 5.1). Comparing the different

curves in Fig. 5.3 in the range of $0.1 \leq d \leq 0.4$, a severe degradation due to phase errors becomes visible. Nevertheless, the optimal value of ρ is still very high for small phase errors ($\gamma = 6$ dB). Even for moderate phase errors with $\gamma = 0$ dB, a small amount of beamforming is meaningful, however, the amount superposing new information becomes dominant because it leads to a higher sum rate. Finally, for severe phase errors with $\gamma = -10$ dB, beamforming does not pay off anymore and should be switched off completely ($\rho = 0$). Increasing d decreases the impact of phase errors due to the decreasing ρ . In particular, even without phase errors, it is superior to use more power for transmission of \tilde{x}_{S2} and thus, the amount used for beamforming (prone to phase error) becomes less. For $d > 0.5$, phase errors have almost no impact on the performance since a high amount of beamforming is not reasonable even without phase errors. As a conclusion, it becomes clear that non-orthogonal MAC is still in the presence of phase errors superior to orthogonal MAC: Small phase errors cause already a significant loss in the achievable rate whereby a substantial amount of the residual gain versus orthogonal channel access is still originated in the array gain due to beamforming. For severe phase errors however, non-orthogonal channel access does only pay off due to the additional virtual transmission from source to relay which is accompanied by additional power in the system (cf. Sec. 3.5.2).³

5.5.2 Simulation Results for Bit Interleaved Coded Modulation Scheme

In contrary to Sec. 5.5.1, the results in this section consider practical coding according to the BICM scheme introduced in Ch. 2. The following subsections show achievable rates as described in Sec. 5.4 versus the SNR on the direct link for different relay positions d and loop SNRs γ . Please note that although 40 MCSs can be exploited, the range of rates is still limited. Hence, feasible solutions for a practical system are only obtained for a limited range of relay distances d and SNRs. Especially $d < 0.2$ would require much higher rates on the source-relay link than available. In order to compare non-orthogonal and orthogonal MAC fairly, results are normalized with respect to the sum transmit energy as in (3.34d) and (3.34g). Hence, the SNR on the direct link is defined as the ratio of received signal energy at destination D to the power spectral noise density N_0 which is normalized to unity.

$$\text{SNR}_{\text{non}} = (2 - \tau_{\text{non}}) \cdot P \quad (5.19a)$$

$$\text{SNR}_{\text{orth}} = P \quad (5.19b)$$

³There are no results for the sum energy constraint because this would require an exhaustive search with respect to parameters δ , β , and τ (cf. (3.35)) which is too complex due to numerical evaluation of the MFMC theorem (cf. App. C.2).

Relay Very Close to Source

Based on the rate allocation in Sec. 5.4, Fig. 5.4 shows the achievable rates for a scenario where R is quite close to D ($d = 0.2$). Hence, distributed beamforming is optimal, that is, $\rho = 1$ for the complete depicted range. The achievable rates obtained according to the MFMC theorem for QAM input and perfect synchronization are plotted additionally as a reference. Obviously, these rates cannot be achieved by a practical code with finite length and only discrete rates. By comparison of the non-orthogonal (solid blue) and the orthogonal (dashed blue) scheme without phase errors, a gain of about 2 dB and more becomes visible. Taking the differently strong impacts of phase errors into account reveals that small phase errors ($\gamma = 10, 5$ dB, cf. Fig. 5.2(a)) barely degrade the performance of the non-orthogonal scheme. Even for moderate phase errors ($\gamma = 0$ dB, cf. Fig. 5.2(b)), the non-orthogonal scheme is still profitable. So far the results of Fig. 5.4 very well confirm the insights of Fig. 5.3. Only for very severe phase errors ($\gamma = -5, -10$ dB, cf. Fig. 5.2(c)) the rate falls below the rate of the orthogonal scheme.⁴ The reason is that $\rho = 1$ is not optimal anymore but the only choice because the SNR on the direct link is too low to allocate even the smallest MCS for a transmission of \tilde{x}_{S2} . Given that, beamforming remains the only choice for the non-orthogonal scheme. As sum energy is considered, the orthogonal scheme would be the better alternative in the presence of severe phase errors.

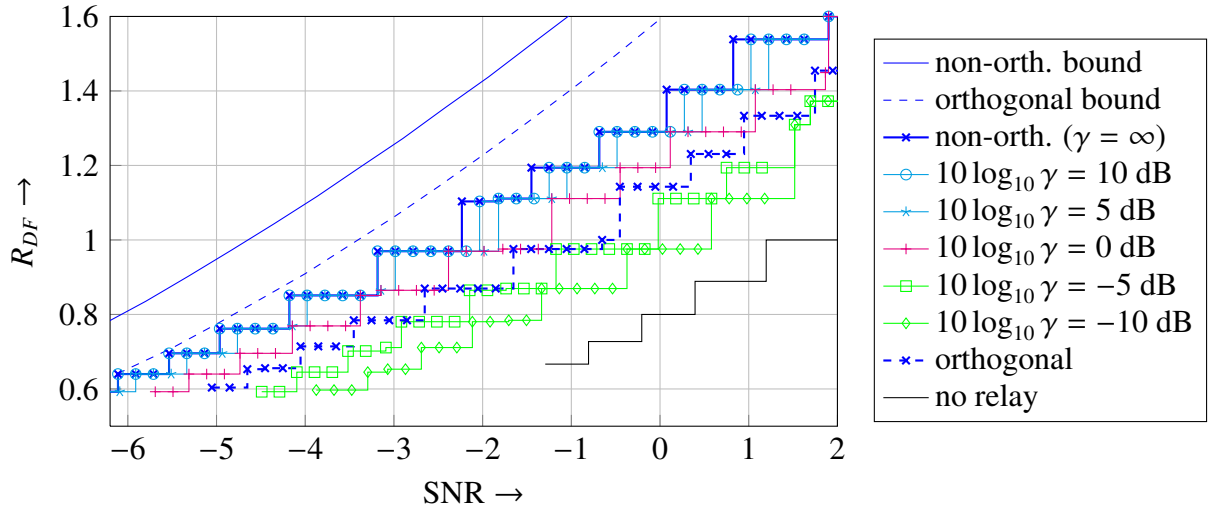


Figure 5.4: Achievable DF rates versus SNR (sum energy) at $d = 0.2$ with $\alpha = 4$ including imperfect phase knowledge dependent on γ . Distributed beamforming with $\rho = 1$ is optimal in the whole range.

⁴It should be emphasized that the Gaussian approximation for the pdf of the phase error is less tight for small values of γ (cf. Fig. 5.2(b) and Fig. 5.2(c)). In particular, the impact of these phase errors should be even more severe.

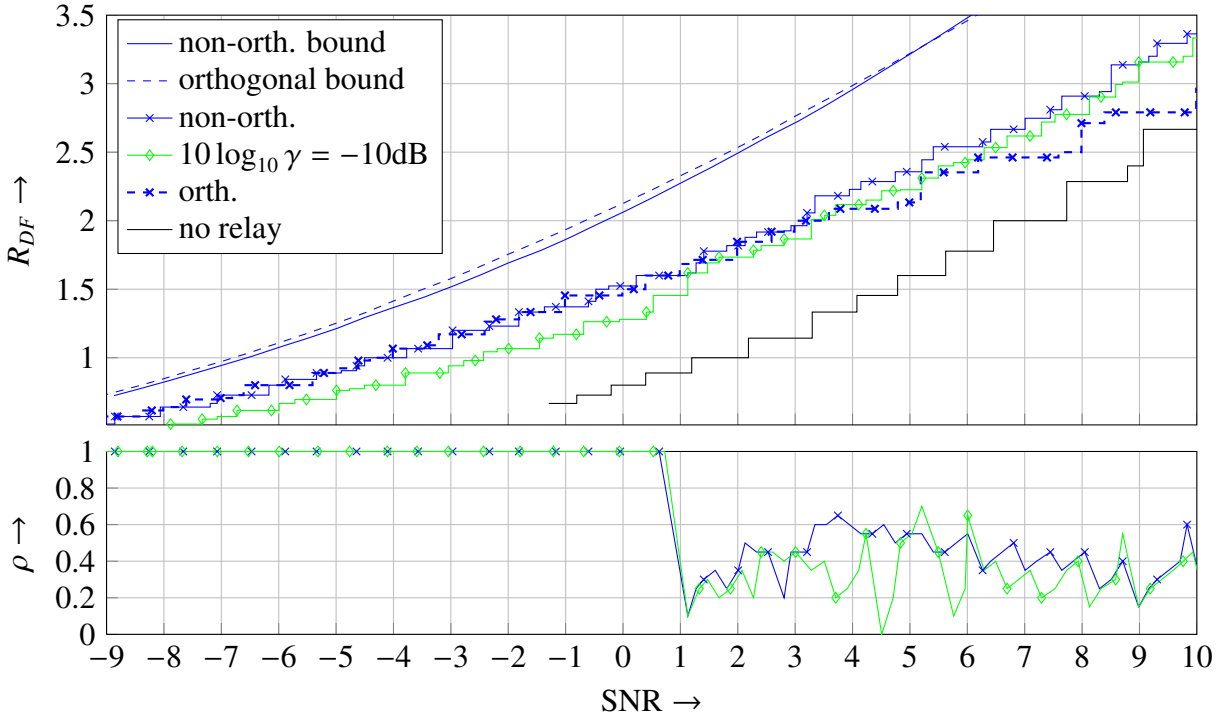
Relay Moderately Close to Source

Figure 5.5: Achievable DF rates versus SNR (sum energy) at $d = 0.4$ with $\alpha = 4$ including imperfect phase knowledge dependent on γ .

Fig. 5.5 shows similar results as before except that the relay is moved a little away from the source to $d = 0.4$. Hence, one could expect generally higher rates which is indeed the case and that the source should superimpose new information ($\rho < 1$) which is only the case in the high SNR regime. The reason is again the limited range of available MCSs to transmit \tilde{x}_{S2} . Hence, not until $\text{SNR} = 1$ dB the non-orthogonal (solid blue) scheme becomes superior than the orthogonal (dashed blue). Before that point, both schemes perform very similar because beamforming is less profitable due to the decreased array gain which is highest for equally strong links.⁵ The oscillating ρ is reasoned by the discreteness of the rates R_b due to the limited number of MCSs: In fact, there are ranges of ρ leading to the same R_{DF} . Anyhow, as ρ is mainly much bigger than zero, applying partly distributed beamforming is still meaningful. Taking phase errors into account, an already very severe case is plotted since the amount of distributed beamforming is much less. Obviously, in the range where $\rho = 1$, the performance is bad. Anyhow, for high SNR, a small amount of beamforming is still applied whereby the performance is better than that of the orthogonal scheme.⁶

⁵Comparing with respect to average power per node instead of sum energy, the non-orthogonal scheme would always be superior.

⁶Due to the fact that some range of ρ delivers the same R_{DF} and only the highest value of that range is plotted, beamforming may be totally switched off for some SNRs.

5.6 Discussion

The influence of phase errors on distributed beamforming for DF has been investigated. Therefore, a random variable is modeled by a first order PLL for which an analytic expression for the respective pdf is known. In a statistical sense, the impact of phase errors is determined by the variance which depends directly on the SNR of the PLL. In particular, if this loop SNR tends to infinity, the probability of the presence of such phase errors goes to zero, and vice versa, for an infinitely low loop SNR, phase errors are uniformly distributed. Using the pdf of the phase error, information theoretic results according to the MFMC have been appropriately extended as shown in Fig. 5.3. Furthermore, simulation results for a BICM system, as described in Sec. 2.3.5 and Sec. 2.4, are obtained for different scenarios. In both cases, it is shown how phase errors degrade the achievable rate of DF for non-orthogonal channel access. Although, small phase errors have a significant impact, there is still a high array gain due to distributed beamforming for which most of the transmit power at the source is still used. Even for moderate phase errors, a small amount of the transmit power is used for beamforming. Only for severe phase errors, beamforming should be switched off completely. In addition to the information theoretic results, simulation results show that superposition coding, is only meaningful when there is a suitable MCS to meet the SNR of the direct link. Unlike in Ch. 6 and Ch. 7 where quantizer design for CF gets more complicated when considering frequency selective channels, DF can be easily applied to an OFDM system with Rayleigh Fading subcarriers. The insights of this chapter will still hold in a relative manner, except that huge phase errors will destroy carrier orthogonality.

Chapter 6

Compress and Forward with Optimized Quantization

In this chapter, practical implementation issues of Compress-and-Forward (CF) are analyzed, more precisely, how to design quantization and evaluation of its performance together with a practical coded modulation scheme like BICM (cf. Sec. 2.3.5). Results have been partly published in [KK16] and are fully shown in Sec. 6.3. Similar results for the multiple access relay channel can be found in [ZKBW08,ZKBW09a,ZKBW09b,ZBW12] where a more sophisticated signal processing is assumed at the relay. The question how to realize and optimize quantization is not straightforward since entropy maximizing quantizers are unrewarding due to dispensable noise which is added at the relay. A suitable answer is given by the IB method [TPB99,Slo02] which maximizes the mutual information (relevant information) between the transmit signal and the quantizer output while compressing the received signal such that its mutual information shared with the transmit signal matches the capacity of the $R \rightarrow D$ link. This concept is furthermore adapted such that the information, conveyed via $S \rightarrow D$, is exploited as side information during compression [Zei12,ZBBW10]. Fig. 6.1 sketches the digital signal processing chain at the relay, where \tilde{y}_R denotes the analog signal. For mathematical description and conducted simulations, y_R is directly addressed with an Analog to Digital (A/D) resolution l_{ad} according to alphabet \mathcal{Y}_R with cardinality $|\mathcal{Y}_R| = 2^{l_{ad}}$.

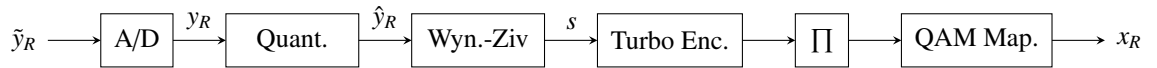


Figure 6.1: Relay Processing for CF.

Sec. 6.1 firstly introduces the IB method and its application to the 3-node relay channel. Afterwards, Sec. 6.2 shows the modified simulation setup including the processing of compression indices in a joint turbo decoder as well as the rate allocation to determine the achievable CF

rate. At the end, Sec. 6.3 shows results in terms of achievable CF rate before Sec. 6.4 concludes this chapter.

6.1 Information Bottleneck Method

The IB method was firstly introduced by Tishby [TPB99, Slo02] and is briefly reviewed in the following subsection. Afterwards, it is adopted to the relay channel as in [ZKBW09a, ZBW12, Zei12] to consider side information which is available at the destination via the direct link in the first time slot.

6.1.1 Information Rate Function

According to Tishby, the quantization of an observation $y \in \mathcal{Y}$ of $x \in \mathcal{X}$ to a compression index $z \in \mathcal{Z}$ can be straightforwardly described by a stochastic mapping $\Pr\{z|y\}$ whereby the random variables $X \rightarrow Y \rightarrow Z$ form a Markov chain. The arising optimization problem is to find $\Pr\{z|y\}$ such that the so called relevant information $I(X; Z)$ is maximized while the source coding rate $I(Y; Z)$ is bounded by r .¹ Mathematically this problem can be defined as an information-rate function [TPB99, Zei12]

$$I(r) \triangleq \max_{\Pr\{z|y\}} I(X; Z) \quad \text{s.t.} \quad I(Y; Z) \leq r \quad (6.1)$$

for $0 < r \leq H(Y)$ or as rate-information function

$$r(I) \triangleq \min_{\Pr\{z|y\}} I(Y; Z) \quad \text{s.t.} \quad I(X; Z) \geq I \quad (6.2)$$

whereby (6.2) is the inverse function, i.e., minimizing the compression rate while a minimum of relevant information is preserved. By the help of Lagrangian multiplier $\beta \geq 0$, which steers the trade-off, the problem may be rewritten as [TPB99, Zei12]

$$r(I(\beta)) - \beta I(\beta) = \min_{\Pr\{z|y\}} I(Y; Z) - \beta I(X; Z). \quad (6.3)$$

An exemplary information-rate curve with slope $1/\beta$ is given in Fig. 6.2. Please note although the information-rate curve is concave, solving (6.3) for each β is a non-convex problem. There are several iterative algorithms [TPB99, Slo02, Zei12] very similar to the Blahut-Arimoto algorithm [Bla72] yielding a local optimum of (6.3). Important to note is that the implementa-

¹The term 'source coding rate' is always used in the context of compression at the relay. Therefore, also the term 'compression rate' will be used meaning the resulting rate $I(Y; Z)$ and not the factor by which has been compressed.

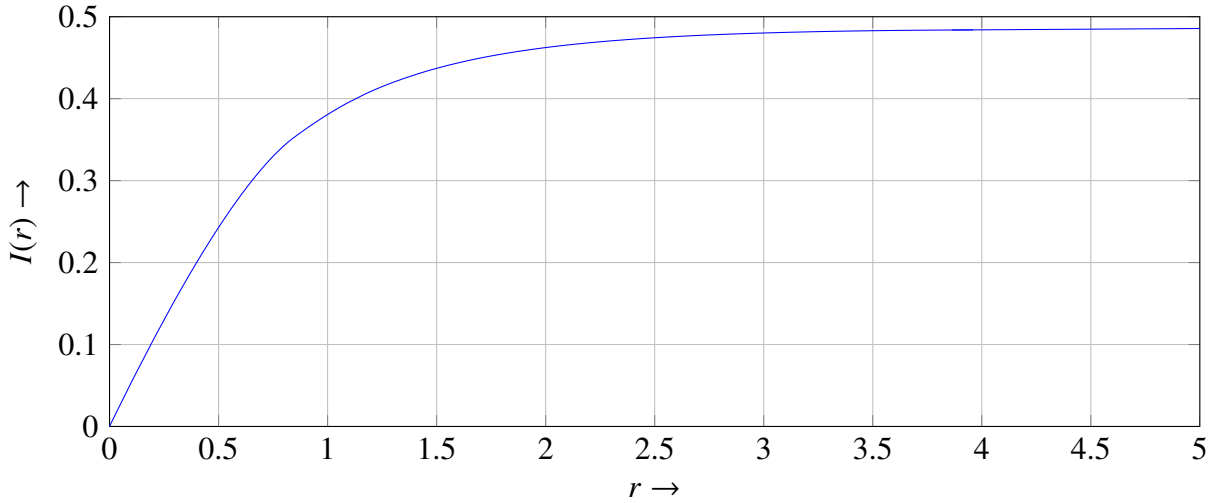


Figure 6.2: Exemplary information-rate curve for $x \in \mathcal{X} = \{-1, 1\}$, $y = x + n$, $n \in \mathcal{N}(0, 1)$, $|\mathcal{Z}| = 64$ and $\epsilon = 10^{-4}$.

tion of these algorithms generally requires discrete random variables described by probability mass functions (pmfs). Hence, for continuous random variables, probability density functions (pdfs) need to be sampled with a sufficiently high resolution into a pmf which yields a quasi-continuous distribution.²

Iterative Information Bottleneck Algorithm

The original algorithm of Tishby, also known as iterative IB algorithm, is depicted in Algorithm 1 [TPB99, Zei12]. As input it takes the joint pmf $\Pr\{x, y\}$, the alphabet \mathcal{Z} , and trade-off parameter β . In addition to $\Pr\{z|y\}$, the algorithm outputs both the a posteriori probability $\Pr\{x|z\}$ and the pair of relevant information $I(\beta)$ and source coding rate $r(\beta)$. More precisely, a specific information-rate pair cannot be directly chosen but is implicitly obtained by choosing a particular value for β . Hence, to calculate the whole information-rate curve as in Fig. 6.2, β has to be varied in the range $0 \leq \beta \leq \infty$. Usually, the aim is not to determine the whole information-rate curve but only one point $(I(r), r)$. Considering (6.1), the rate r of a specific point $(I(r), r)$ is given while $I(r)$, $\Pr\{z|y\}$, and $\Pr\{x|z\}$ are wanted. As $(I(r(\beta)), r(\beta))$ depends on β , a bisection search has to be applied to meet, e.g., a given rate r . The initialization of $\Pr\{z|y\}$ is usually random [TPB99, Zei12]. However, due to the non-convex nature of the problem, random initialization cannot guarantee to find a global optimum. Hence, typically several runs are necessary to find a close to optimum value for $(I(r(\beta)), r(\beta))$. To avoid several runs, Maximum Output Entropy (MOE) instead of random initialization is used throughout this work [Win14]. As shown

²This happens inherently in the digital system under investigation due to A/D conversion which acts as a uniform pre-quantization.

Input: $\Pr\{x, y\}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}, \beta > 0, \epsilon > 0$

Output: $\Pr\{z|y\}, \Pr\{x|z\}, (I(r(\beta)), r(\beta))$

```

1 initialize  $\Pr\{z|y\}^{(0)}$ 

2  $\Pr\{z\}^{(0)} \leftarrow \sum_y \Pr\{y\} \Pr\{z|y\}^{(0)}$ 
3  $\Pr\{x|z\}^{(0)} \leftarrow \frac{1}{\Pr\{z\}^{(0)}} \sum_y \Pr\{x, y\} \Pr\{z|y\}^{(0)}$ 
4  $d^{(0)}(z, y) \leftarrow D_{KL}(\Pr\{x|y\} \| \Pr\{x|z\}^{(0)})$ 
5  $\Pr\{z|y\}^{(1)} \leftarrow \frac{\Pr\{z\}^{(0)} 2^{-\beta d^{(0)}(z, y)}}{\sum_{z^*} \Pr\{z^*\}^{(0)} 2^{-\beta d^{(0)}(z^*, y)}}$ 
6  $k \leftarrow 1$ 

7 while  $\sum_{z \in \mathcal{Z}, y \in \mathcal{Y}} |\Pr\{z|y\}^{(k)} - \Pr\{z|y\}^{(k-1)}| / (|\mathcal{Y}| \cdot |\mathcal{Z}|) \geq \epsilon$  do
8    $\Pr\{z\}^{(k)} \leftarrow \sum_y \Pr\{y\} \Pr\{z|y\}^{(k)}$ 
9    $\Pr\{x|z\}^{(k)} \leftarrow \frac{1}{\Pr\{z\}^{(k)}} \sum_y \Pr\{x, y\} \Pr\{z|y\}^{(k)}$ 
10   $d^{(k)}(z, y) \leftarrow D_{KL}(\Pr\{x|y\} \| \Pr\{x|z\}^{(k)})$ 
11   $\Pr\{z|y\}^{(k+1)} \leftarrow \frac{\Pr\{z\}^{(k)} 2^{-\beta d^{(k)}(z, y)}}{\sum_{z^*} \Pr\{z^*\}^{(k)} 2^{-\beta d^{(k)}(z^*, y)}}$ 
12   $k \leftarrow k + 1$ 
13 end

14  $\Pr\{z|y\} \leftarrow \Pr\{z|y\}^{(k)}$ 
15  $\Pr\{z\} \leftarrow \sum_y \Pr\{y\} \Pr\{z|y\}$ 
16  $\Pr\{x|z\} \leftarrow \frac{1}{\Pr\{z\}} \sum_y \Pr\{x, y\} \Pr\{z|y\}$ 
17  $I(r(\beta)) \leftarrow \sum_{x, z} \Pr\{x|z\} \Pr\{z\} \log_2 \left( \frac{\Pr\{x|z\}}{\Pr\{x\}} \right)$ 
18  $r(\beta) \leftarrow \sum_{y, z} \Pr\{z|y\} \Pr\{y\} \log_2 \left( \frac{\Pr\{z|y\}}{\Pr\{z\}} \right)$ 

```

Algorithm 1: Iterative IB algorithm [TPB99], where D_{KL} denotes relative entropy (cf. App. A.5) or Kullback-Leibler-Divergence [CT91].

in [Win14], MOE initialization yields only suboptimal but still good results. Specifying the cardinality $|\mathcal{Z}|$ of \mathcal{Z} is not only important for initialization, it also limits the compression rate $r \leq H(Z)$ indirectly via its entropy. Hence, it has to be ensured that $|\mathcal{Z}|$ is sufficiently large to cover the whole information-rate curve. Therefore, Fig. 6.3 shows information-rate curves for differing cardinalities $|\mathcal{Z}_1| = 2$, $|\mathcal{Z}_2| = 3$, $|\mathcal{Z}_3| = 4$, and $|\mathcal{Z}_4| = 16$ whereby the most right points correspond to $\beta \rightarrow \infty$.

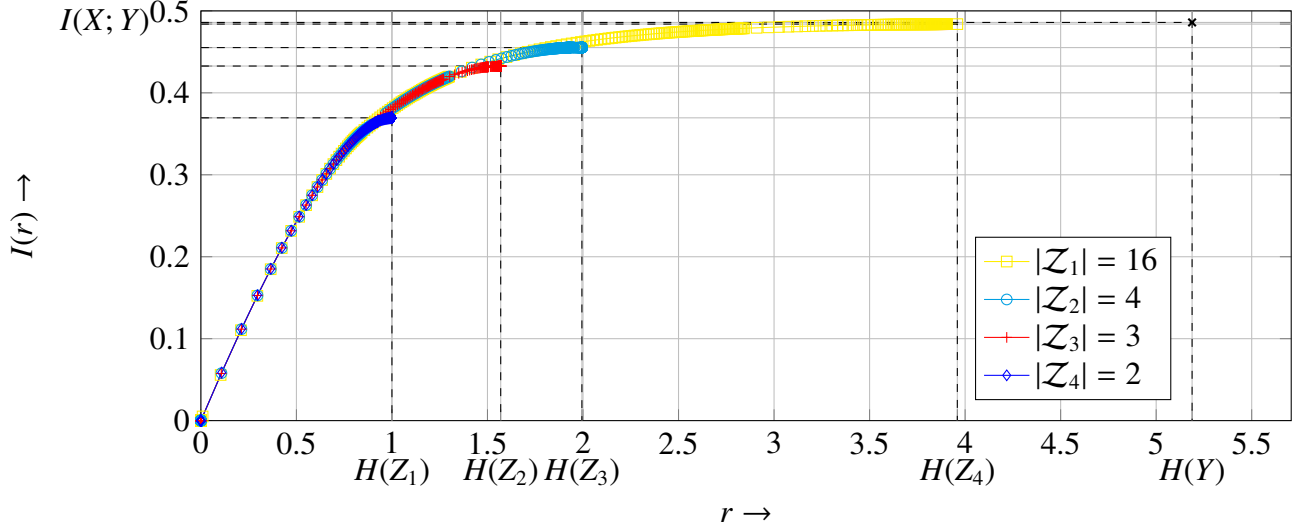


Figure 6.3: Information-rate curve for different cardinalities $|\mathcal{Z}_1| = 2$, $|\mathcal{Z}_2| = 3$, $|\mathcal{Z}_3| = 4$, and $|\mathcal{Z}_4| = 16$ with $x \in \mathcal{X} = \{-1, 1\}$, $y = x + n$, $n \in \mathcal{N}(0, 1)$, and $\epsilon = 10^{-4}$.

Iterative Information Bottleneck Algorithm for $\beta \rightarrow \infty$ (KL-Means Algorithm)

The problem in (6.3) represents the trade-off between a minimum quantizer rate $I(Y; Z)$ and maximum relevant information $I(X; Z)$ which cannot be achieved at the same time but is steered by β . The extreme cases are $\beta = 0$ and $\beta \rightarrow \infty$ whereby the first case is not of interest because it minimizes the rate $I(Y; Z)$ and thus also the relevant information $I(X; Z)$ to zero. The other extreme however has a practical relevance as it maximizes rate $I(Y; Z)$ and relevant information $I(X; Z)$ for a given given \mathcal{Z} [Kur17]. This leads to a deterministic quantizer as maximizing $I(Y; Z) = H(Z) - H(Z|Y)$ implies $H(Z|Y) = 0$, that is, $\Pr\{z|y\} \in \{0, 1\}$.³ Hence, the resulting mapping $\Pr\{z|y\}$ can be implemented as a Look-Up-Table (LUT). Due to $\beta \rightarrow \infty$, the trade-off between relevant information and compression rate depends only on the cardinality $|\mathcal{Z}|$ since (6.3) is solved for a specific \mathcal{Z} . In particular, the maximum rate r for a specific \mathcal{Z} is its corresponding entropy $H(Z)$. As the cardinality is integer, there is only a discrete (instead of the beforehand continuous) range of rates r . This effect is very well depicted in Fig. 6.3 where

³Generally, for arbitrary $\beta > 0$, the IB method yields a random mapping $\Pr\{z|y\} \in [0, 1]$.

only the most right point of each information-rate curve corresponds to a solution of (6.3) if $\beta \rightarrow \infty$. Fig. 6.3 shows furthermore that although the relevant information is maximized for a specific cardinality $|\mathcal{Z}|$ it is slightly less than that of a comparable information-rate pair with higher $|\mathcal{Z}|$ and variable $\beta \in [0, \infty]$. Balancing pros and cons, the only considerable drawback of setting $\beta \rightarrow \infty$ is the discreteness of the range of rates $r = H(Z)$ depending on $|\mathcal{Z}|$. It will be explained in Sec. 6.1.2 that this limited range of rates is not disadvantageous with respect to the simulation setup in Sec. 6.2 to determine the achievable rate for CF. Finally, Algorithm 1 is modified such that $\beta \rightarrow \infty$ yields Algorithm 3 in App. F. In particular, line 5 in Algorithm 1 simply changes to $\Pr\{\hat{y}_R|y_R\}^{(1)} \leftarrow \mathbb{I}_{\hat{y}_R=\hat{y}_R^*}$ with $\hat{y}_R^* = \operatorname{argmin}_{\hat{y}_R} d^{(0)}(\hat{y}_R, y_R)$, $\forall y_R$ where

$$\mathbb{I}_{\hat{y}_R=\hat{y}_R^*} := \begin{cases} 1, & \text{if } \hat{y}_R = \hat{y}_R^* \\ 0, & \text{otherwise} \end{cases} \quad (6.4)$$

denotes the indicator function.

6.1.2 Relay Channel Adapted Information Rate Function

Considering the 3-node relay channel for CF according to Fig. 6.4 and the corresponding achievable rate in (3.17), (6.1) can be extended to the trade-off between $I(X_{S1}; \hat{Y}_R|Y_{D1})$ and $I(Y_R; \hat{Y}_R|Y_{D1})$ exploiting the side information which is available at destination D due to the broadcast phase. More precisely, it is possible to determine the final source coding rate r at the

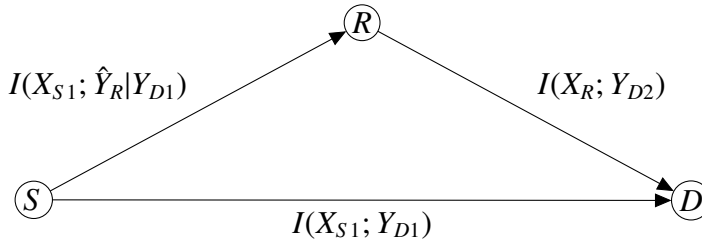


Figure 6.4: 3-node relay channel with CF.

output of a Wyner-Ziv encoder. It is important to note that although r is the rate after Wyner-Ziv coding, the IB method determines the mapping $\Pr\{\hat{y}_R|y_R\}$ and not a mapping $\Pr\{b|y_R\}$ to a bin index b according to random binning as in App. B. In other words, the rate $I(Y_R; \hat{Y}_R)$ before Wyner-Ziv source coding is usually higher than the rate $I(Y_R; \hat{Y}_R|Y_{D1})$ afterwards. Following the derivation in [Zei12], the extended information rate function given the joint distribution $\Pr\{x_{S1}, y_R, y_{D1}\}$ is defined as

$$I(r) \triangleq \max_{\Pr\{\hat{y}_R|y_R\}} I(X_{S1}; \hat{Y}_R|Y_{D1}) \quad \text{s.t.} \quad I(Y_R; \hat{Y}_R|Y_{D1}) \leq r, \quad (6.5)$$

where $0 < r \leq H(Y_R|Y_{D1})$ denotes the rate after source coding with side information (Wyner-Ziv). Applying the method of Lagrangian multipliers as in Sec. 6.1, (6.5) may be rewritten as

$$r(I(\beta)) - \beta I(\beta) = \min_{\Pr\{\hat{y}_R|y_R\}} I(Y_R; \hat{Y}_R|Y_{D1}) - \beta I(X_{S1}; \hat{Y}_R|Y_{D1}). \quad (6.6)$$

Relay Channel Adapted Iterative Information Bottleneck Algorithm

From (6.3) the modified iterative IB algorithm, as shown in Algorithm 2, can be derived [Zei12]. This algorithm delivers the mapping $\Pr\{\hat{y}_R|y_R\}$, the statistics $\Pr\{x_{S1}|\hat{y}_R\}$, and the information rate pair $(I(r(\beta)), r(\beta))$ for a given input distribution $\Pr\{x_{S1}, y_R, y_{D1}\}$ for a specific β . According to the CF relay protocol, r is restricted by the capacity of the relay destination link such that $\tau \cdot r \leq (1 - \tau) \cdot I(X_R; Y_{D2})$ holds (cf. (3.19)). Obviously, simplification to deterministic quantization by setting $\beta \rightarrow \infty$ is possible similar as before (cf. Algorithm 4 in App. F). The input pmf $\Pr\{x_{S1}, y_R, y_{D1}\}$ for the adapted iterative IB algorithm is given in the following. According to (3.15a), (3.15b) and x_{S1} being a uniformly distributed QAM symbol, all random variables can be split into their respective real and imaginary parts denoted by a prime and a double prime. Hence, I and Q component of y_R can be quantized independently with the same quantizer because $\Pr\{x_{S1}, y_R, y_{D1}\}$ is symmetric in real and imaginary part. The real joint pmf is

$$\Pr\{x'_{S1}, y'_R, y'_{D1}\} = \Pr\{y'_R|x'_{S1}\} \Pr\{y'_{D1}|x'_{S1}\} \Pr\{x'_{S1}\} \quad (6.7)$$

where $\Pr\{y'_R|x'_{S1}\}$ and $\Pr\{y'_{D1}|x'_{S1}\}$ are obtained by appropriate discretization of Gaussian distributions according to $\mathcal{N}(\sqrt{P_{S1}}h_{SR}x'_{S1}, \frac{1}{2})$ and $\mathcal{N}(\sqrt{P_{S1}}h_{SD}x'_{S1}, \frac{1}{2})$. The resolution of these pmfs is defined by the A/D converter with resolution l_{ad} which is set to $l_{ad} = 8$ bit for the simulation in Sec. 6.2. Given $\Pr\{x'_{S1}, y'_R, y'_{D1}\}$, Algorithm 2 (and also Algorithm 4) delivers the mapping $\Pr\{\hat{y}'_R|y'_R\}$ and the statistics $\Pr\{x'_{S1}|\hat{y}'_R\}$ which will be exploited in a joint decoder at the destination.

Extension to Multivariate Compression

It is important to note that all depicted algorithms like Algorithm 2 and Algorithm 1 can generally be applied to multivariate variables. Therefore, the multivariate variables (described by a vector) need to be stacked into one dimension [LSB16] to obtain a pmf of the form $\Pr\{x_{S1}, y_R, y_{D1}\}$. Then, the first dimension generally corresponds to relevant variable(s), the second to quantizer input variable(s) (e.g. observed variable(s)), and the third to side information. In the next chapter, Information Bottleneck Graphs (IBGs) will be used to graphically emphasize this structure.

Input: $\Pr\{x_{S1}, y_R, y_{D1}\}, \mathcal{X}_{S1}, \mathcal{Y}_R, \mathcal{Y}_{D1}, \hat{\mathcal{Y}}_R, \beta > 0, \epsilon > 0$

Output: $\Pr\{\hat{y}_R|y_R\}, \Pr\{x_{S1}|\hat{y}_R\}, (I(r(\beta)), r(\beta))$

```

1 initialize  $\Pr\{\hat{y}_R|y_R\}^{(0)}$ 

2  $k \leftarrow 1$ 
3  $\Pr\{\hat{y}_R\}^{(0)} \leftarrow \sum_{y_R} \Pr\{y_R\} \Pr\{\hat{y}_R|y_R\}^{(0)}$ 
4  $\Pr\{\hat{y}_R, y_{D1}\}^{(0)} \leftarrow \sum_{y_R} \Pr\{y_R, y_{D1}\} \Pr\{\hat{y}_R|y_R\}^{(0)}$ 
5  $\Pr\{\hat{y}_R|y_{D1}\}^{(0)} \leftarrow \frac{\Pr\{\hat{y}_R, y_{D1}\}^{(0)}}{\Pr\{y_{D1}\}}$ 
6  $\Pr\{x_{S1}|\hat{y}_R, y_{D1}\}^{(0)} \leftarrow \frac{1}{\Pr\{\hat{y}_R, y_{D1}\}^{(0)}} \sum_{y_R} \Pr\{x_{S1}, y_R, y_{D1}\} \Pr\{\hat{y}_R|y_R\}^{(0)}$ 
7  $d^{(0)}(\hat{y}_R, y_R) \leftarrow \beta \sum_{y_{D1}} \Pr\{y_{D1}|y_R\}$ 
    $D_{KL}(\Pr\{x_{S1}|y_R, y_{D1}\} || \Pr\{x_{S1}|\hat{y}_R, y_{D1}\}^{(0)}) - \sum_{y_{D1}} \Pr\{y_{D1}|y_R\} \log_2(\Pr\{\hat{y}_R|y_{D1}\}^{(0)})$ 
8  $\Pr\{\hat{y}_R|y_R\}^{(1)} \leftarrow 2^{-d^{(0)}(\hat{y}_R, y_R)} / \sum_{\hat{y}_R^*} 2^{-d^{(0)}(\hat{y}_R^*, y_R)}$ 

9 while  $\sum_{\hat{y}_R \in \hat{\mathcal{Y}}_R, y_R \in \mathcal{Y}_R} |\Pr\{\hat{y}_R|y_R\}^{(k)} - \Pr\{\hat{y}_R|y_R\}^{(k-1)}| / (|\mathcal{Y}_R| \cdot |\hat{\mathcal{Y}}_R|) \geq \epsilon$  do
10    $\Pr\{\hat{y}_R\}^{(k)} \leftarrow \sum_{y_R} \Pr\{y_R\} \Pr\{\hat{y}_R|y_R\}^{(k)}$ 
11    $\Pr\{\hat{y}_R, y_{D1}\}^{(k)} \leftarrow \sum_{y_R} \Pr\{y_R, y_{D1}\} \Pr\{\hat{y}_R|y_R\}^{(k)}$ 
12    $\Pr\{\hat{y}_R|y_{D1}\}^{(k)} \leftarrow \frac{\Pr\{\hat{y}_R, y_{D1}\}^{(k)}}{\Pr\{y_{D1}\}}$ 
13    $\Pr\{x_{S1}|\hat{y}_R, y_{D1}\}^{(k)} \leftarrow \frac{1}{\Pr\{\hat{y}_R, y_{D1}\}^{(k)}} \sum_{y_R} \Pr\{x_{S1}, y_R, y_{D1}\} \Pr\{\hat{y}_R|y_R\}^{(k)}$ 
14    $d^{(k)}(\hat{y}_R, y_R) \leftarrow \beta \sum_{y_{D1}} \Pr\{y_{D1}|y_R\}$ 
    $D_{KL}(\Pr\{x_{S1}|y_R, y_{D1}\} || \Pr\{x_{S1}|\hat{y}_R, y_{D1}\}^{(k)}) - \sum_{y_{D1}} \Pr\{y_{D1}|y_R\} \log_2(\Pr\{\hat{y}_R|y_{D1}\}^{(k)})$ 
15    $\Pr\{\hat{y}_R|y_R\}^{(k+1)} \leftarrow 2^{-d^{(k)}(\hat{y}_R, y_R)} / \sum_{\hat{y}_R^*} 2^{-d^{(k)}(\hat{y}_R^*, y_R)}$ 
16    $k \leftarrow k + 1$ 
17 end

18  $\Pr\{\hat{y}_R|y_R\} \leftarrow \Pr\{\hat{y}_R|y_R\}^{(k)}$ 
19  $r(\beta) \leftarrow \sum_{y_R, \hat{y}_R} \Pr\{\hat{y}_R|y_R\} \Pr\{y_R\} \log_2\left(\frac{\Pr\{\hat{y}_R|y_R\}}{\Pr\{\hat{y}_R\}}\right) - \sum_{y_{D1}, \hat{y}_R} \Pr\{\hat{y}_R|y_{D1}\} \Pr\{y_{D1}\} \log_2\left(\frac{\Pr\{\hat{y}_R|y_{D1}\}}{\Pr\{\hat{y}_R\}}\right)$ 
20  $I(r(\beta)) \leftarrow \sum_{x_{S1}, y_{D1}, \hat{y}_R} \Pr\{x_{S1}|\hat{y}_R, y_{D1}\} \Pr\{\hat{y}_R, y_{D1}\} \log_2\left(\frac{\Pr\{x_{S1}|\hat{y}_R, y_{D1}\}}{\Pr\{x_{S1}|y_{D1}\}}\right)$ 

```

Algorithm 2: Iterative IB algorithm considering side information [Zei12].

Deterministic versus Random Quantization

In Sec. 6.1 is shown that $\beta \rightarrow \infty$ simplifies the general IB problem such that the quantizer mapping $\Pr\{\hat{y}_R|y_R\}$ is deterministic and can be implemented as LUT. The trade-off between relevant information $I(X_{S1}; \hat{Y}_R|Y_{D1})$ and compression rate $r = I(Y_R; \hat{Y}_R|Y_{D1}) = H(\hat{Y}_R|Y_{D1})$ is then steered by the cardinality $|\hat{\mathcal{Y}}_R|$ of the compressed variable \hat{y}_R . Therefore, the range of available compression rates r is now discrete due to the integer nature of cardinality. This however is not a problem for determining a suitable quantizer for CF in the 3-node relay channel such that $\tau \cdot r = (1 - \tau) \cdot I(X_R; Y_{D2})$. In particular, as neither the optimal r nor the optimal τ can be obtained in closed form, an exhaustive search including a high effort Monte Carlo simulation is necessary (cf. Sec. 6.2). Generally, it does not matter whether a grid of τ or r is defined: The respective other appears due to $\tau \cdot r = (1 - \tau) \cdot I(X_R; Y_{D2})$. However, if deterministic quantization is used, a set of rates $r = H(\hat{Y}_R|Y_{D1})$ is inherently defined by a set of different cardinalities $|\hat{\mathcal{Y}}_R|$. Hence, the limited range due to setting $\beta \rightarrow \infty$ is no drawback at all but an actual benefit because Algorithm 4 has to be run only once for a specific rate indirectly determined by $|\hat{\mathcal{Y}}_R|$.⁴ On the contrary, Algorithm 2 (determining a random quantizer mapping) would have to run several times such that a predefined rate r is met within a bisection search. Especially for low values of r and $I(r)$, the number of runs within this search increases significantly due to the almost linear shape of the information-rate curve (cf. Fig. 6.2 or Fig. 6.3). The reason is that, the search intervals are usually reduced by a factor of 2 with a reasonable starting point of, e.g., $\beta = 500$. In the almost linear regime of the information-rate function, the relevant variation of β , which represents the slope of the function, becomes negligibly small (a couple of decades). To visualize this problem, Fig. 6.5 shows different rates r in dependence on the SNR on the direct link of the 3-node relay channel. The different rates r are either obtained by varying $|\hat{\mathcal{Y}}_R| \in \{2, \dots, 16\}$ for deterministic quantization (red) or by an equidistant grid $r \in \{\frac{r_{max}}{6}, \dots, r_{max}\}$ with $r_{max} = H(\hat{Y}_R|Y_{D1})$ according to $|\hat{\mathcal{Y}}_R| = 16$ for general random quantization (blue). All red curves are monotonically decreasing because the uncertainty, which defines the rate $r = H(\hat{Y}_R|Y_{D1})$, is lowered with increasing SNR.⁵ In other words, with increasing SNR, i.e., amount of noise in Y_R decreases, a stronger compression is possible while keeping the same amount of relevant information $I(X_{S1}; \hat{Y}_R)$. The observation of general decrease is transferable to the blue curves because of $r_{max} = H(\hat{Y}_R|Y_{D1})$. However, due to the beforehand stated numerical issues of the bisection method, there is a varying error on the targeted rate which is especially high for low rates. Comparing red and blue curves with respect to the covered range, random quantization may only be advantageous for low SNR because it provides very low quantizer rates below 1. Unfortunately, this advantage comes with a very high price, namely, computational complexity. The maximum number of bisection iterations was set to 100,

⁴This approach is suboptimal because this artificially predefined set of rates may not contain the optimal rate r . Anyhow, this is true for any artificial sampling which has to be coarse for the sake of feasibility.

⁵Due to numerical reasons the entropy does not decrease with every step of 0.1.

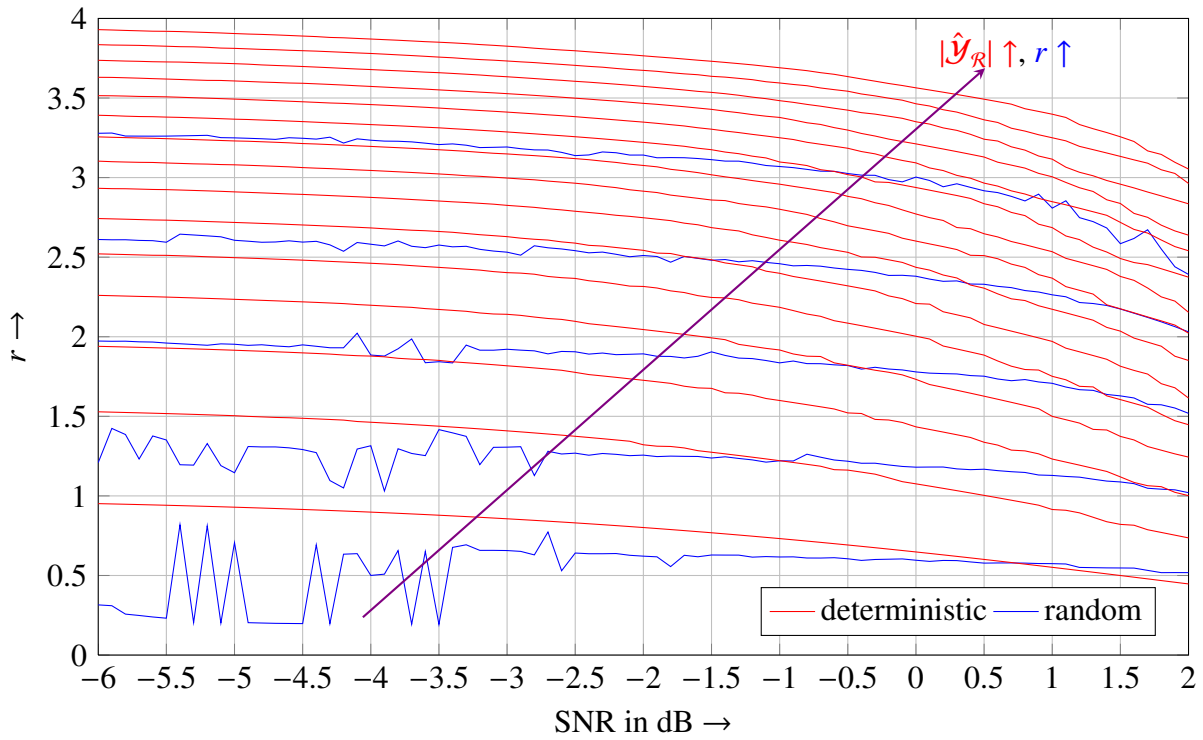


Figure 6.5: Different quantizer rates r versus SNR (direct link of the 3-node relay channel) comparing deterministic quantization ($r = H(\hat{Y}_R|Y_{D1}) \forall |\hat{\mathcal{Y}}_R| \in \{2, \dots, 16\}$) to random quantization ($r \in \{\frac{r_{max}}{6}, \dots, r_{max}\}$) where r_{max} corresponds to $|\hat{\mathcal{Y}}_R| = 16$ (red and blue match).

i.e., in worst case the IB algorithm is carried out 100 times instead of only once. Furthermore, the huge deviations from $H(\hat{Y}_R|Y_{D1})/6$ confirm that still after 100 iterations the targeted rate is not reached. This numerical instability can be caused by different reasons, e.g., the almost linear shape of the information-rate function or suboptimal MOE initialization of $\Pr\{\hat{y}_R|y_R\}$. Due to the just mentioned reasons and the aim of straightforward implementation as LUTs, deterministic quantization will be used throughout this work.

6.2 Modified Simulation Setup for Compress and Forward

As the achievable rate for CF can only be determined considering a decoder which jointly decodes y_{D1} and \hat{y}_R , the simulation setup from Sec. 2.4 has to be extended. Therefore, it is necessary to clarify how the quantizer indices \hat{y}_R are interpreted by a joint turbo decoder. Using both the results of the modified and the beforehand simulation, rate allocation can be done analogue to Sec. 5.4.

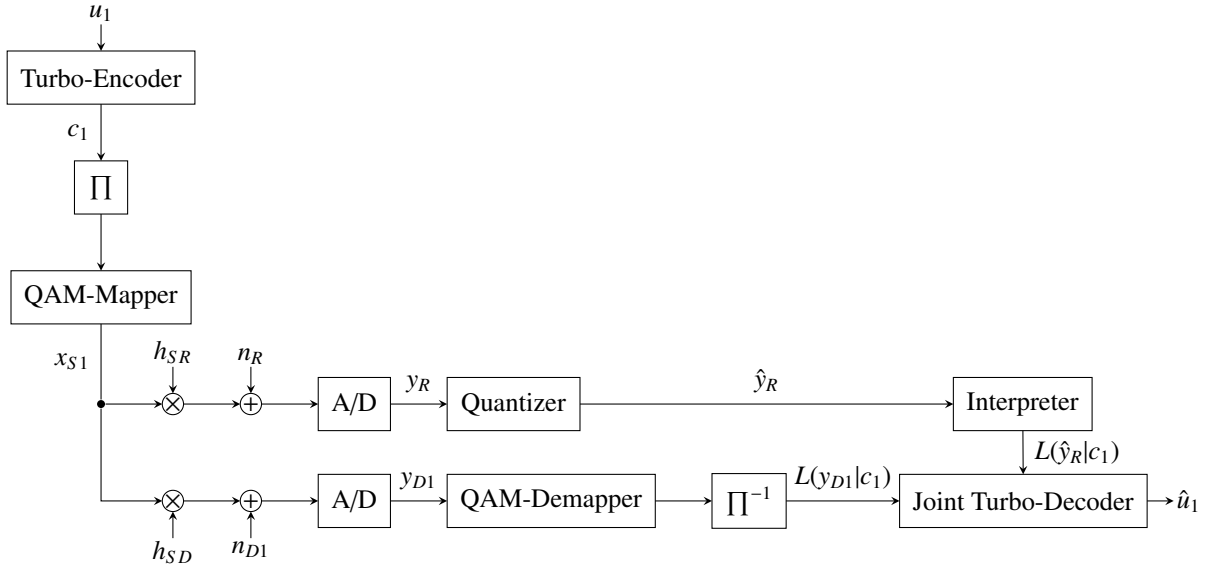


Figure 6.6: Block diagram for CF transmission with joint destination processing.

6.2.1 Processing of Compression Indices in Soft Decoders

Fig. 6.6 shows the simulation setup to determine the achievable CF rate for the BICM scheme introduced in Sec. 2.3.5 (cf. Fig. 2.21). The depicted setup does not include the MAC but only the broadcast phase to keep the complexity of the simulation feasible. In fact, it is assumed that the relay can forward \hat{y}_R with a specific rate R_b^{xR} in the MAC phase which is determined as explained in Sec. 2.4. Thus, Fig. 6.6 does not show the channel and the required signal processing between the quantizer output at the relay and the interpreter at the destination. Both

the quantizer and the interpreter are designed as LUTs: The first is simply the mapping $\Pr\{\hat{y}_R|y_R\}$ prior obtained by the IB method. The second maps \hat{y}_R to an LLR $L(\hat{y}_R|c_1)$ with respect to the code bits c_1 which can be calculated a priori from $\Pr\{x_{S1}|\hat{y}_R\}$ (also determined by the IB method). Please note that real and imaginary parts are independently quantized with the same quantizer, that is, $\Pr\{\hat{y}'_R|y'_R\} = \Pr\{\hat{y}''_R|y''_R\}$ and $\Pr\{x'_{S1}|\hat{y}'_R\} = \Pr\{x''_{S1}|\hat{y}''_R\}$. The variables \hat{y}_R and $L(\hat{y}_R|c_1)$ in Fig. 6.6 represent multiplexed versions of their respective real and imaginary parts which implies that the same interpreter LUT can be used to convert \hat{y}'_R and \hat{y}''_R into an LLR.

$$\begin{aligned}
L(\hat{y}'_R|c_1) &= \log \frac{\Pr\{\hat{y}'_R|c_1 = 0\}}{\Pr\{\hat{y}'_R|c_1 = 1\}} \\
&= \log \frac{\sum_{x_{S1}} \Pr\{\hat{y}'_R|x'_{S1}\} \Pr\{x'_{S1}|c_1 = 0\}}{\sum_{x_{S1}} \Pr\{\hat{y}'_R|x'_{S1}\} \Pr\{x'_{S1}|c_1 = 1\}} \\
&= \log \frac{\sum_{x_{S1}} \Pr\{x'_{S1}|\hat{y}'_R\} \Pr\{x'_{S1}|c_1 = 0\} / \Pr\{x'_{S1}\}}{\sum_{x_{S1}} \Pr\{x'_{S1}|\hat{y}'_R\} \Pr\{x'_{S1}|c_1 = 1\} / \Pr\{x'_{S1}\}} \quad (6.8)
\end{aligned}$$

The component decoders of the joint turbo decoder in Fig. 6.6 are usual BCJR decoders and the exchanging of extrinsic LLRs is as in Sec. 2.3.2. The only difference is that the decoder takes the sum $L(y'_{D1}|c) + L(\hat{y}'_R|c)$ as input instead of only $L(y'_{D1}|c)$ (cf. App. G). This seems natural since \hat{y}'_R and y'_{D1} are independent observations of x'_{S1} .

6.2.2 Rate Allocation

The above presented simulation setup will be used, similar as in Sec. 2.4, to obtain BERs versus SNR (direct link) for each available Modulation and Coding Scheme (MCS).⁶ In addition, the simulation is carried out for a set of different quantizers with compression rates r according to $|\hat{\mathbf{y}}_R| \in \{2, \dots, 16\}$ as in Fig. 6.5. The reason is that neither τ nor r can be optimally determined a priori (cf. (3.19)). Hence, a set of BERs is obtained for each MCS which yields a set of achievable rates $R_b^{xs1}(r)$ for a specific threshold SNR at a target BER of 10^{-5} . Fig. 6.7 shows the same rates as in Fig. 2.22 for an AWGN channel and additionally the rates $R_b^{xs1}(r)$ for the setup depicted in Fig. 6.6 whereby only the achievable rates $R_b^{xs1}(r)$ for one specific quantizer ($|\hat{\mathbf{y}}_R| = 8$) are plotted. The quantizer rate versus SNR on the direct link is depicted in Fig. 6.5. The gap between the respective curves in Fig. 6.7 visualizes the gain of a CF transmission over a direct transmission. This gap increases or decreases with the source coding rate r of the quantizer at the relay which depends on the achievable rate $R_b^{xr}_{(non)}$ (non-orthogonal MAC) or $R_b^{xr}_{(orth)}$ (orthogonal MAC) on the relay-destination link and τ . Hence, to determine the

⁶As complexity (cf. Sec. 6.2.3) is an issue, the maximum simulated modulation order is $m = 3$ instead of $m = 5$ as in Sec. 2.4. All other parameters like interleaver length or the number of statistical realizations is like in Sec. 2.4

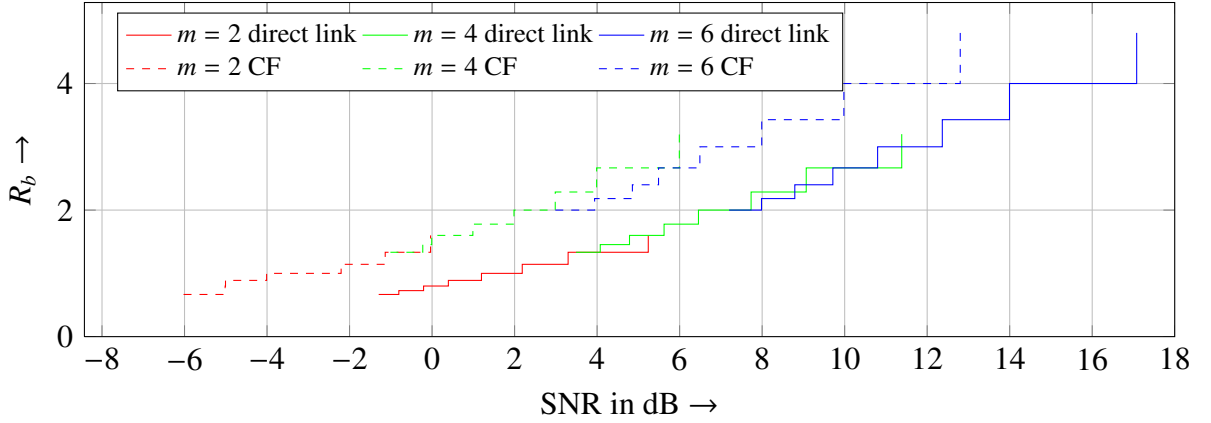


Figure 6.7: Achievable rates R_b^{XS1} versus SNR on direct $S \rightarrow D$ link for $|\hat{\mathcal{Y}}_R| = 8$ and $d = 0.8$ compared to achievable rates for a direct transmission. Different colors indicate different modulation orders.

finally achievable rate for CF, R_b^{XR} and R_b^{XS2} (only for non-orthogonal MAC) are determined as in Sec. 5.4.⁷ The final CF rate is

$$R_{CF}^{orth} = \max_r \{ \tau_{orth} R_b^{XS1}(r) \} \quad \text{s.t.} \quad \tau_{orth} \cdot r \leq (1 - \tau_{orth}) R_b^{XR}_{(orth)} \quad (6.9a)$$

$$R_{CF}^{non} = \max_r \{ \tau_{non} R_b^{XS1}(r) + (1 - \tau_{non}) R_b^{XS2} \} \quad \text{s.t.} \quad \tau_{non} \cdot r \leq (1 - \tau_{non}) R_b^{XR}_{(non)} \quad (6.9b)$$

whereby the maximization is over all simulated quantizers with respective compression rates r according to $|\hat{\mathcal{Y}}_R| \in \{2, \dots, 16\}$.

6.2.3 Complexity

In the previous chapter, the complexity was fully determined by the Monte Carlo simulations to find threshold SNRs for the predefined MCSs given a specific target BER (cf. Fig. 2.22). The effort to calculate the final rates for DF (cf. Sec. 5.4) is comparably negligible. For CF however, the complexity is significantly higher because the optimal values for τ and r cannot be determined a priori and, thus, different combinations are tested whereby only one parameter has to be varied since the other follows from $\tau \cdot r = (1 - \tau) \cdot R_b^{XR}$. Hence, the Monte Carlo simulation has to be carried out not only for each MCS and over a range of SNRs but also for a range of, e.g., r . Although the simulation setup in Sec. 2.4 differs slightly from the one in Sec. 6.2, the effort for one simulation over a range of SNRs is almost the same. For DF, all required individual rates R_b^{XS1} , $R_b^{XR}_{(non)}$, $R_b^{XR}_{(orth)}$, and R_b^{XS2} are determined based on the simulations in Sec. 2.4 which are 40 (one for each MCSs over a range of SNRs). For CF however, the rate $R_b^{XS1}(r)$ is not determined by the setup in Sec. 2.4 but by the setup in Sec. 6.2 whereby $R_b^{XR}_{(non)}$,

⁷Please note that R_b^{XS2} for CF is similar to R_b^{XS2} for DF. Both denote the rate of the superposed and independently encoded new information word u_2 which is transmitted by the source in the MAC phase (cf. Ch. 3).

$R_{b \text{ (orth)}}^{xR}$, and R_b^{xS2} are determined likewise as for DF. Hence, 360 additional simulations are required: For each of the 24 MCSs (modulation order limited to $m = 3$), 15 different quantizers are used in the simulations. Finally, the 15 different quantizers have to be determined once for each SNR and each modulation order m which adds a small but not negligible amount of complexity.

6.3 Achievable Rates for Compress and Forward with Optimized Quantization

Similar as for DF in Sec. 5.5.2, the following results show achievable rates for CF versus the SNR on the direct link whereby all nodes are placed on a line with individual transmit powers $P_{S1} = P_{S2} = P_R = P$. The achievable rates R_{CF} are compared to the rates R_{DF} of DF. Furthermore, the results are analyzed with respect to orthogonal and non-orthogonal MAC whereby both the individual power constraint and the sum energy constraint are considered (cf. Sec. 3.5). The results in this section are updated and more detailed than in [KK16].

6.3.1 Comparison to Decode and Forward

In the following, the achievable rate of CF will be compared to the achievable rate of DF and to that of a direct transmission which does not use the relay. Both non-orthogonal and orthogonal MAC are considered.

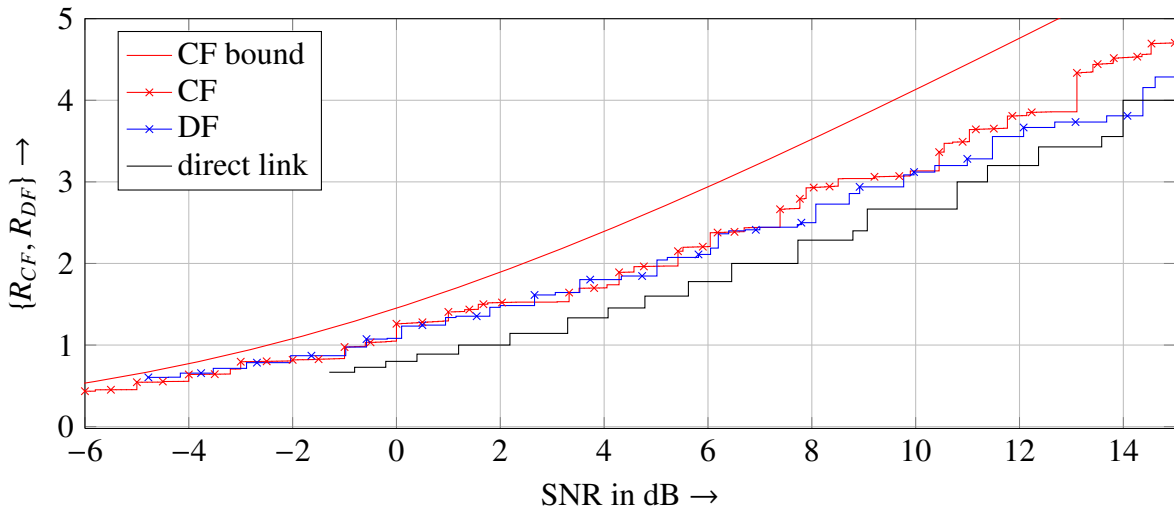


Figure 6.8: Total achievable rates for CF and DF versus sum SNR (sum energy) at $d = 0.8$ for non-orthogonal channel access in the MAC phase.

Non-Orthogonal Channel Access

To compare the performance of CF and DF, Fig. 6.8 shows the achievable rates for a relay position $d = 0.8$ whereby the MCSs are automatically adapted according to (6.9) and (5.18). Almost in the whole depicted range DF is outperformed by CF. For some SNRs however, DF performs equally good or better. When the relay is moved closer to the destination as in Fig. 6.9, CF clearly outperforms DF. The reason is, as mentioned before, that the $S \rightarrow R$ link becomes the bottleneck of the system. In addition to the achievable rates for DF and CF, Figures 6.8 and 6.9 show the achievable rates for direct transmission without using the relay. Obviously, CF is always superior while DF does not pay off for $d = 0.9$ in the high SNR regime.

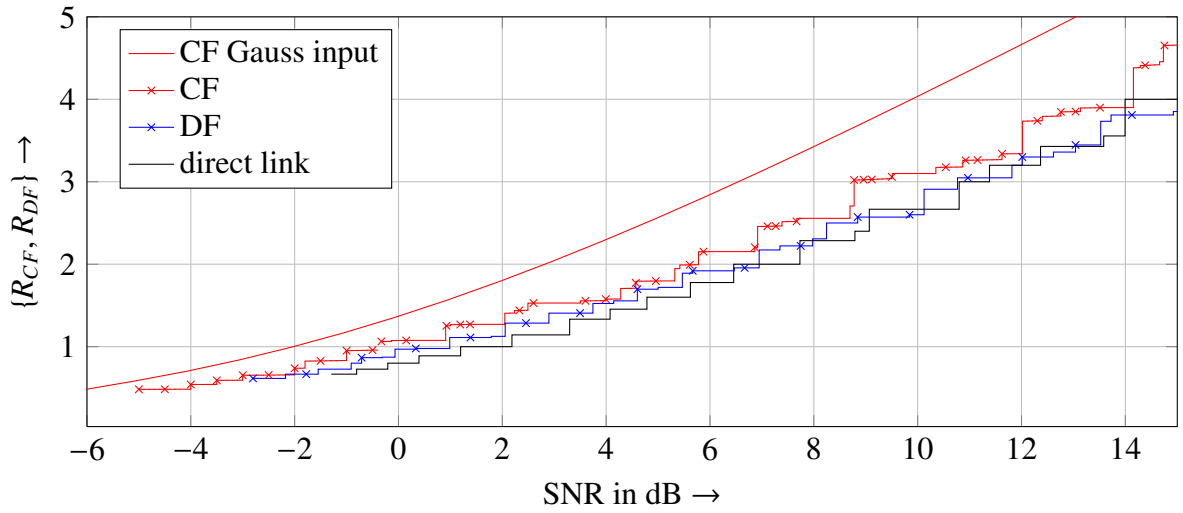


Figure 6.9: Total achievable rates for CF and DF versus SNR (sum energy) at $d = 0.9$ for non-orthogonal channel access in the MAC phase.

Orthogonal Channel Access

Comparing CF and DF for orthogonal MAC in the second time slot as shown in Fig. 6.10 and Fig. 6.11, the same relative behavior as before can be observed. Anyhow, in contrary to DF, CF clearly outperforms the direct transmission (no relay): For $d = 0.8$ in Fig. 6.10, CF is superior in the whole range while DF is only beneficial up to 6 dB. For $d = 0.9$ in Fig. 6.11, DF has only a slight gain in the low SNR regime while CF is still superior up to the mid SNR regime. The reason for the inferior performance of CF versus the direct link in the high SNR regime is reasoned by the saturation on the $R \rightarrow D$ link due to the limited range of MCSs.

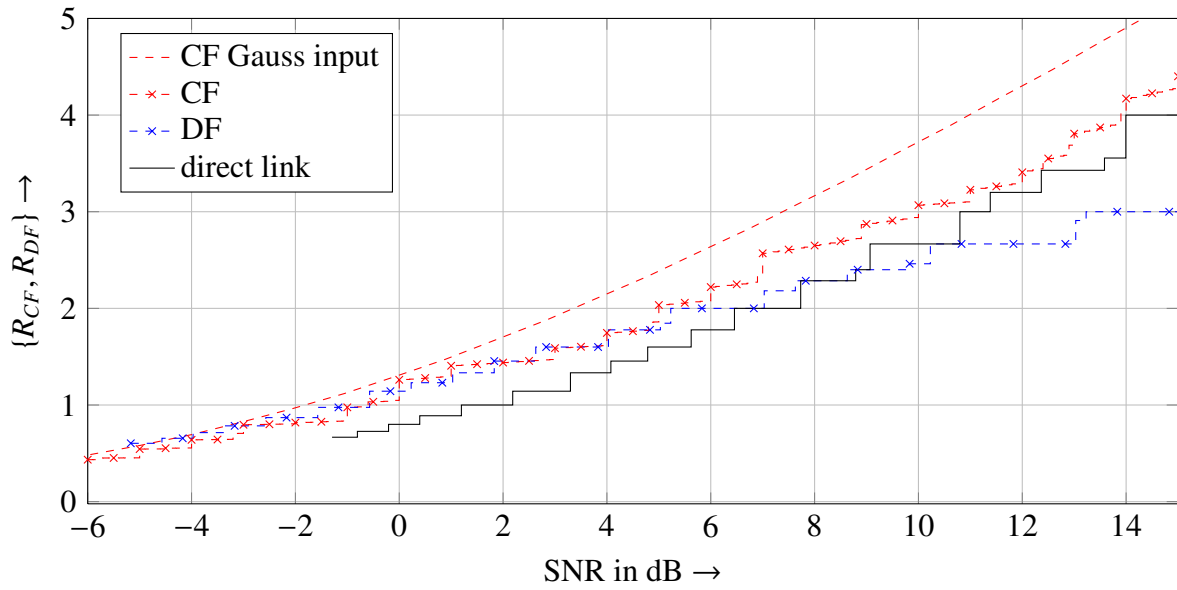


Figure 6.10: Total achievable rates for CF and DF versus SNR (sum energy) at $d = 0.8$ for orthogonal channel access in the MAC phase.

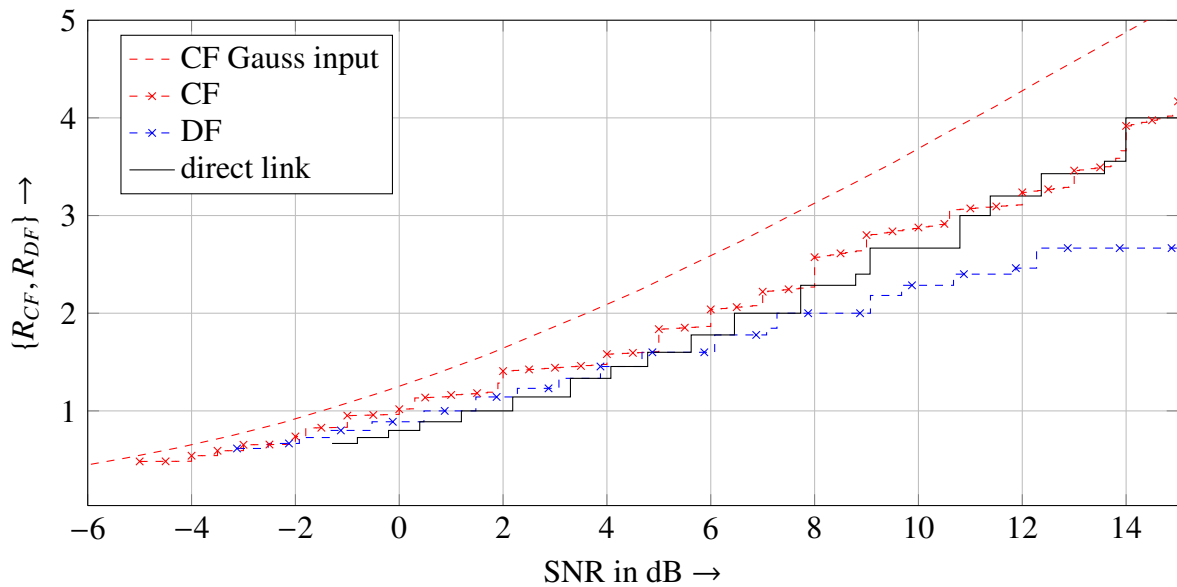


Figure 6.11: Total achievable rates for CF and DF versus SNR (sum energy) at $d = 0.9$ for orthogonal channel access in the MAC phase.

6.3.2 Comparison of Non-Orthogonal and Orthogonal Scheme

Achievable Rate for Sum Energy

The comparison of non-orthogonal and orthogonal channel access shows similar results for CF and DF (cf. Sec. 5.5). Fig. 6.12 shows the respective rates R_{CF} for $d = 0.8$. In the low SNR range, both curves coincide up to about 0 dB. The reason is the low capacity of the $S \rightarrow D$ link since even the smallest MCS has a too high rate to ensure sufficient decoding. Hence, the source remains quiet in the MAC phase leading to orthogonal channel access.

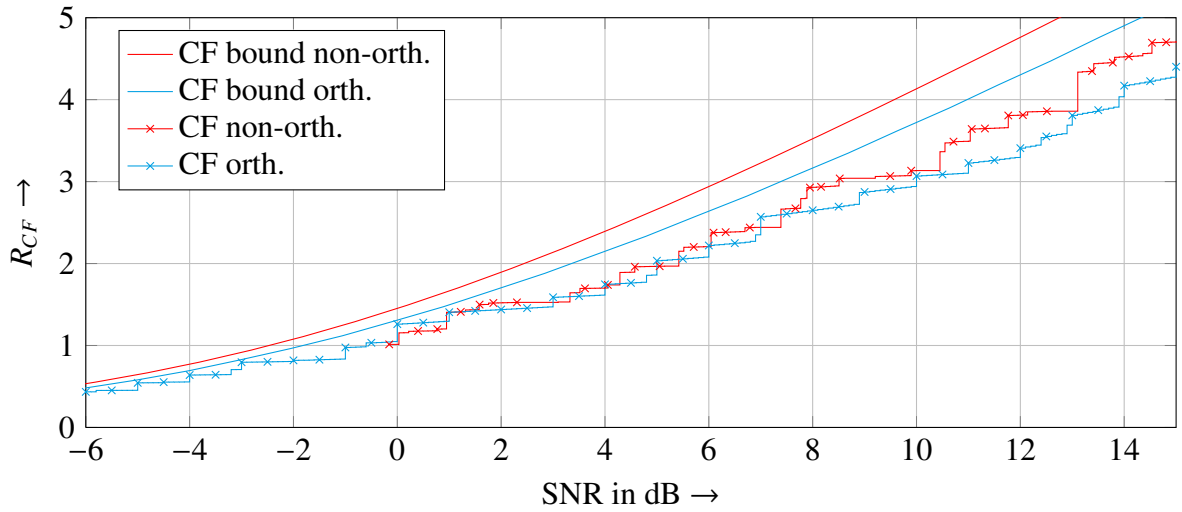


Figure 6.12: Total achievable rates for CF versus SNR (sum energy) at $d = 0.8$ comparing orthogonal to non-orthogonal channel access in the MAC phase.

In the moderate SNR range, both schemes perform very similar so that the additional effort of non-orthogonal channel access becomes unrewarding. In fact, this insight is only true with respect to the sum energy constraint. For another measure of comparison, e.g., individual transmit power as in Fig. 6.13, the non-orthogonal scheme will always outperform the orthogonal if an MCS can be assigned to the direct link between S and D . For high SNRs in Fig. 6.12, the orthogonal scheme is clearly outperformed by the non-orthogonal scheme: On one hand, the non-orthogonal scheme has higher degrees of freedom in terms of time and rate allocation (which will be shown in the following in Fig. 6.13). On the other hand, the $R \rightarrow D$ link is limited by the highest practically available MCS that is 8 bit/s/Hz in the illustrated setup.⁸ This saturation of the $R \rightarrow D$ link degrades the orthogonal scheme more than the non-orthogonal because the “weak” $S \rightarrow D$ can be used for transmission with higher MCSs in the high SNR regime. Using this link in both time slots (non-orthogonal) is now highly beneficial compared to using it only in the first time slot (orthogonal). In other words, without the limitation due to

⁸Please note that this MCS corresponds to a 1024-QAM with code rate 4/5 which is at the limit of state-of-the-art implementation.

practical MCS, that is, if the $R \rightarrow D$ capacity could be achieved, the orthogonal scheme would perform slightly better and the gap to the non-orthogonal scheme would be smaller.

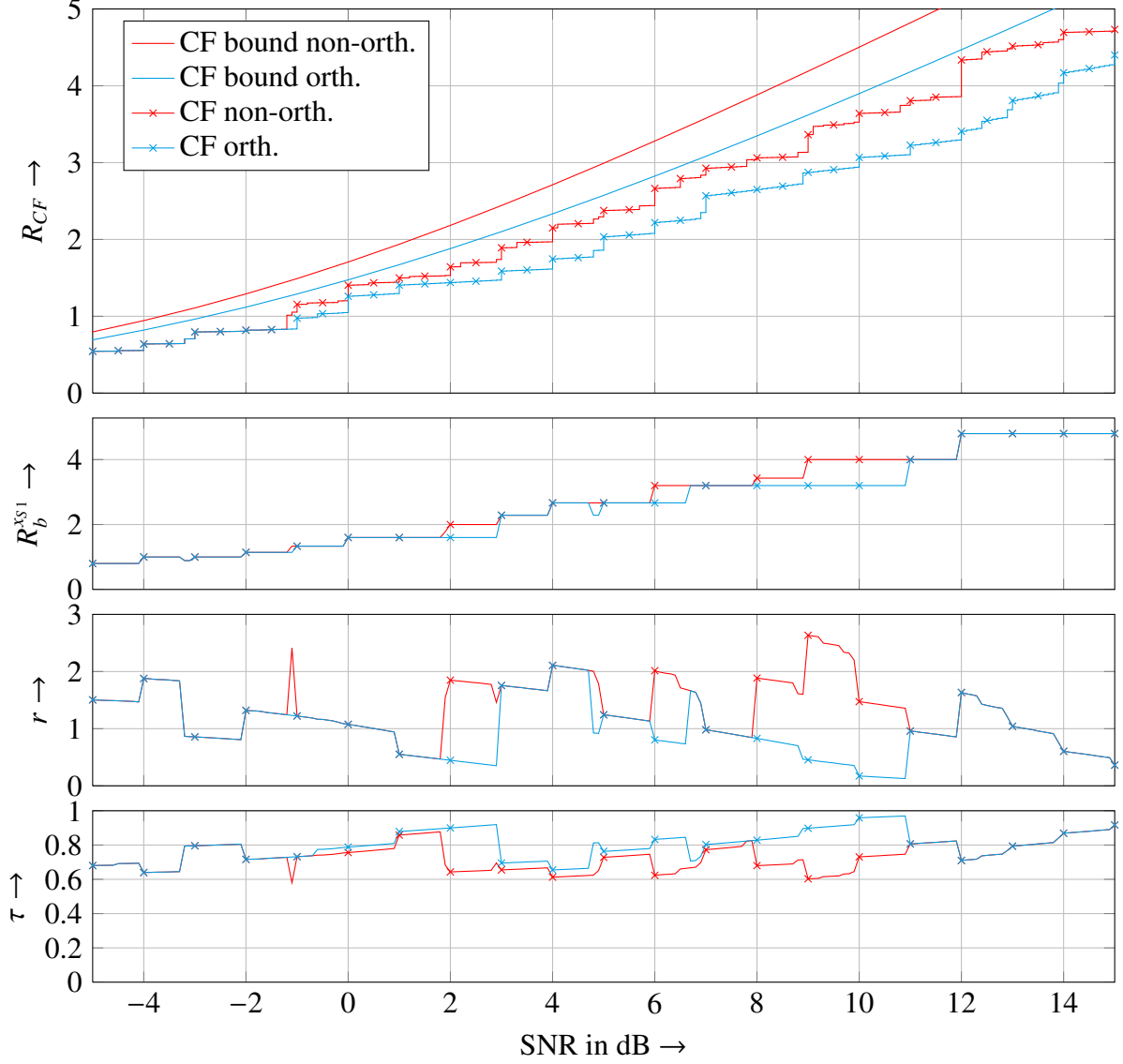


Figure 6.13: Total achievable rates for CF versus SNR (individual power) at $d = 0.8$ comparing orthogonal to non-orthogonal channel access in the MAC phase.

Achievable Rate for Individual Power per Node

As before, Fig. 6.13 compares the performance of non-orthogonal and orthogonal channel access for CF whereby here, in contrary to Fig. 6.12, achievable rate is plotted with respect to transmit power per node. Additionally, Fig. 6.13 shows the respective parameters $R_b^{xs1}(r)$, r , and τ after optimization with respect to R_{CF} according to (6.9). The general behavior of the rates R_{CF}^{non} and R_{CF}^{orth} is a step wise increase with the SNR which is originated in the discrete nature

of the MCSs (similar as for DF) as well as in the artificial discreteness of r (cf. Fig. 6.5). Despite this step size nature both R_{CF}^{non} and R_{CF}^{orth} increase as well within each step which becomes especially visible for the orthogonal scheme and is reasoned in the decreasing behavior of r (increasing τ) within one step size: Such a behavior is meaningful because an increasing SNR allows stronger compression to reduce r while keeping the same relevant information particularly because r is determined by $H(\hat{Y}_R|Y_{D1})$ depending on $|\hat{\mathcal{Y}}_R|$. According to (6.9), the decrease of r allows an increase of τ . This effect is not that obvious for the non-orthogonal scheme due to the addition of R_b^{xs2} which increases only step-by-step whenever a threshold SNR for a higher MCS is exceeded. From the behavior of the partial parameters, it is obvious that the discreteness of r and the MCSs (R_b^{xs1} , R_b^{xR} , R_b^{xs2}) significantly reduces the degrees of freedom in determining the optimal rate R_{CF} . The general trend seems to be a decreasing r and an increasing τ especially when the highest MCS for R_b^{xs1} is reached. Anyhow, due to only limited discrete availability of transmission and source coding rates, the best trade-off does not follow a monotone behavior but jumps unpredictably. As the non-orthogonal schemes have one degree of freedom more, due to transmission via R_b^{xs2} , the best trade-off between τ , r , and R_b^{xs1} differs in some points from that of the orthogonal scheme.

6.4 Discussion

In this chapter, the IB has been used to determine optimal quantizers to apply CF to the 3-node relay channel as introduced in Sec. 3.1. As general random quantization may cause numerical problems, the iterative IB is modified to deliver deterministic quantizers. This modification simplifies on one hand the complexity of the algorithm but on the other hand allows only a limited range of source coding rates which are determined by the cardinality of the compressed variable instead of a Lagrangian multiplier. This includes also the fact that the source coding rate cannot be arbitrarily small which is generally possible. Nevertheless, for the problem at hand, a limited number of source coding rates is not a drawback for the sake of feasibility. It rather is a benefit because it is much easier to generate a meaningful range of source coding rates in dependence of cardinality instead of a Lagrangian Multiplier which depends on the actual shape of the information-rate function. The suchlike obtained deterministic quantizers are stored in straightforward implementable LUTs and used to determine achievable rates for a BICM system. Simulation results show an expected better performance of CF over DF and a direct transmission when the relay is considerably closer to the destination than to the source. For the single-carrier system and the AWGN channel model with path loss which has been considered so far, feasibility to obtain optimal quantizers is not an issue. However, when extending the channel model to a more realistic frequency selective channel which is tackled via a multi-

carrier scheme like OFDM, quantizer design gets much more complex and suboptimal design approaches have to be found. This topic is addressed in the following chapter.

Chapter 7

Quantizer Design for Compress and Forward over Multiple Carriers

In this chapter, the quantizer design of the previous chapter is generalized to consider frequency selective channels that are decomposed into flat Rayleigh fading channels via OFDM (cf. Sec. 2.5). In contrary to DF, where SNRs and SINRs are simply converted into an effective value, this extension is not straightforward for CF.¹ In fact, the random nature of the channel has to be incorporated into the quantizer design. In theory, there are different possibilities to do that while only a few are feasible in practice. Firstly, it would be desirable to have a low cost relay, that is, to have a minimum amount of signal processing. For CF, it would be theoretically sufficient to directly quantize the received signal and apply a subsequent source and channel coding to forward the quantizer indices. It has already been shown that a quantizer can be implemented as a LUT optimized using the IB method. Furthermore, such an integer based signal processing can be generalized, e.g., for LDPC decoders or equalization with very low complexity [LB15,LSB16]. Unfortunately, quantization in time domain is not feasible due to the channel memory L_h . More precisely, the IB method requires the joint pmf of the channel input (relevant variable), its output (observation), and additionally the channel itself as random variable. Due to the inter-symbol-interference caused by the channel, one input sample of a time domain OFDM symbol is present in a vector of received samples of length $L_h + 1$. Hence, the required pmf would have a cardinality which grows exponentially with the channel length. Except for very minimalist toy examples such a pmf would require a main storage far beyond Terabytes. Therefore, quantization is implemented independently for each of the orthogonal subcarriers in frequency domain where different possibilities arise: The simplest way would be to equalize the subcarriers and do the quantization as in the previous chapter. Unfortunately, this approach is not practical as will be shown in Sec. 7.1. An alternative is to treat the actual channel coefficient

¹The effective SNR or SINR is obtained by converting each value into a mutual information and subsequent averaging over all subcarriers. This average mutual information is then converted back into an effective value.

as quantizer input, that is, quantization is optimized for the ensemble of channel coefficients. Doing so leads to a higher dimensionality of the joint pmfs required by the IB algorithm. Furthermore, the side information for Wyner-Ziv coding should cover channel knowledge which further increases the dimensionality of the respective input pmf. The extra dimensions due to additional random variables yield such a high cardinality that storage of these pmfs becomes infeasible.² As a consequence, this chapter proposes design approaches for suboptimal but feasible quantizers using Information Bottleneck Graphs (IBGs) [LSB16]. These IBGs are used to simplify the dimensionality and the cardinality of the respective pmfs, that is, one infeasible IB problem is decomposed into several feasible problems. During the work on this chapter, three publications have been originated which are concerned with different quantizer design approaches as well as perfect and imperfect channel knowledge [KK17a, KK17b, KK17c]. There is a similar work in the literature which uses a more sophisticated relay to decode the received vector and forward quantized LLRs [DSZ⁺11]. On the contrary, the following presented scheme works without the need of a costly decoder keeping the goal of a relay, which uses solely integer based signal processing, in mind.

The chapter is organized as follows: Sec. 7.1 introduces additional preliminaries like IBGs which are basically factor graphs. Afterwards, Sec. 7.2 and Sec. 7.3 derive suitable quantizer designs for perfect [KK17a, KK17c] and imperfect channel knowledge [KK17b], respectively. Finally, simulation results in terms of achievable rates are presented in Sec. 7.4 before Sec. 7.5 concludes the chapter.

7.1 Preliminaries

As stated above, quantizer design for a state-of-the-art OFDM system is not straightforward: In time domain, quantization is not feasible due to an exponentially growing storage complexity caused by the channel memory. Therefore, quantization will be done in frequency domain which requires an FFT before an independent quantization per subcarrier (cf. Fig. 7.1). The theoretically optimal and intuitive way would be to equalize (cf. Sec. 2.5.2) each carrier and design the quantizers as in the previous chapter for a range of relevant SNRs. Anyhow, such a scheme cannot be practically implemented because the rate of each quantizer depends on the instantaneous channel realization of each subcarrier which determines the SNR. Hence, the rate of an OFDM symbol varies depending on the instantaneous vector channel.³ This scheme is nevertheless theoretically treated in App. H and used as reference in the simulation results. The alternative for practical implementation is to consider the channel coefficient or an equivalent

²The complexity here is huge but still far less complex than a comparable approach in time domain.

³To be honest, general random quantization and setting a specific rate for the IB algorithm, would avoid this problem. However, despite numerical issues, this would increase run time complexity dramatically because a bisection search has to be framed around the IB algorithm for each possible SNR (cf. Sec. 6.1).

like the observed pilot symbol as an additional input to the quantizer (cf. Fig. 7.1). Doing so, this quantizer is valid for the ensemble of channel coefficients (all possible SNRs) and can be applied for all subcarriers which makes its output rate independent from the instantaneous channel realization. Generally, this approach requires neither equalization nor channel estimation which is a first step in the direction of a relay using solely integer based signal processing. Nevertheless, explicit equalization before quantization has benefits as will be shown in Sec. 7.2 and Sec. 7.4. The entire relay processing chain is depicted in Fig. 7.1 whereby quantization

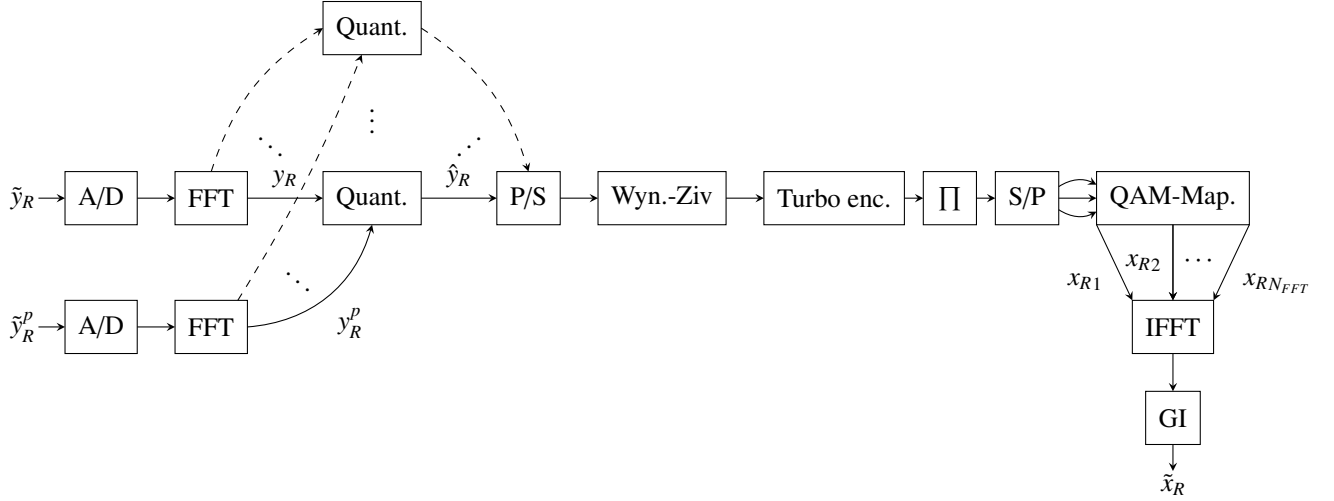


Figure 7.1: Relay Processing for CF with direct quantization of y_R exploiting observed pilot y_R^p .

is highlighted to be in a per subcarrier manner. Therefore, other carriers are only indicated by dashed lines. For the sake of clarity, only the general approach of direct quantization is shown. Anyhow, a second approach using explicit channel estimation and equalization before as introduced in Sec. 2.5.2 will be investigated as well in Sec. 7.2. Just as in Sec. 6.2, Wyner-Ziv source coding and the subsequent processing are only shown for the sake of completeness but are not explicitly simulated. As the focus is on quantizer design, the same holds for the A/D conversion and the FFT which are depicted to emphasize the need for a pre-quantization with an A/D resolution of l_{ad} . To be precise, the analysis of this chapter is completely in the frequency domain, that is, y_R and y_R^p are obtained according to the flat Rayleigh fading channel model in Sec. 2.1.2. Although quantization in frequency domain is applicable independently along the subcarriers with the same quantizer, the required pmf for the optimal design with an IB algorithm would still require too much main storage (cf. Table 7.1). The reason is simply the high number and cardinality of different dimensions in the required pmf especially when side information is considered. Such a pmf and especially its factorization to reduce complexity can be visualized with a factor graph as shown in the following. Afterwards, factor graphs are further extended to obtain an Information Bottleneck Graph (IBG) to highlight IB optimized compress-

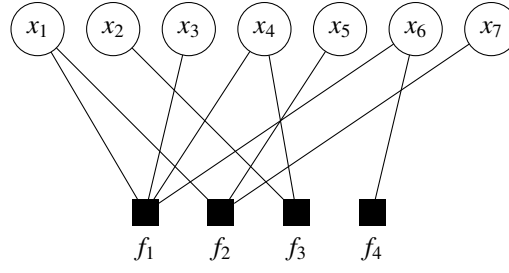


Figure 7.2: Exemplary factor Graph.

sion mappings [LSB16]. It is important to note that throughout the rest of this chapter notation is done for one subcarrier because the resulting quantizer is applicable for all subcarriers.

7.1.1 Factor Graphs

In general, factor graphs are bipartite graphs known to be a graphical presentation of the factorization of a global function. [KFL01, Ksc03]

$$F(v_1, \dots, v_n) = \prod_{j=1}^J f_j(\mathcal{V}_j) \text{ with } \mathcal{V}_j \subseteq \{v_1, \dots, v_n\} \quad (7.1)$$

The function arguments v_l of the global function $F(v_1, \dots, v_n)$ are variable nodes which are usually drawn as circles. Factor nodes, usually depicted as squares, represent the local functions $f_j(\mathcal{V}_j)$ which are only connected to a specific variable node if this node is an argument of the local function f_j . Fig. 7.2 illustrates an exemplary factor graph which represents the factorization of global function $F(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ such that

$$F(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = f_1(x_1, x_3, x_4, x_6) f_2(x_1, x_5, x_7) f_3(x_2, x_4) f_4(x_6). \quad (7.2)$$

In communications, factor graphs are very well known to describe, e.g., LDPC codes to emphasize their decoding by message passing [KFL01, Ksc03]. The factor nodes in such codes on graphs usually represent probabilities or related measures like LLRs that are passed as messages between factor and variable nodes. Revisiting Fig. 7.2 with the assumption of functions being pdfs, one has the possibility to modify the graph to marginalize variables by merging, e.g., variable x_6 with its connected factors f_1 and f_4 leading to the graph in Fig. 7.3 with $f_{14}(x_1, x_3, x_4) = \sum_{x_6 \in \mathcal{X}_6} f_1(x_1, x_3, x_4, x_6) f_4(x_6)$.

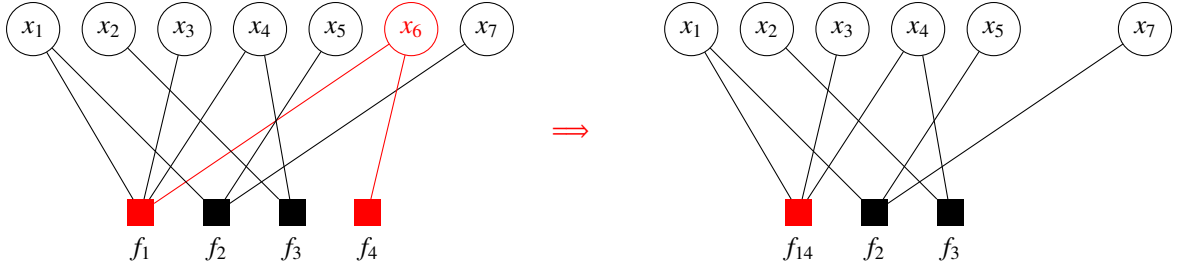


Figure 7.3: Factor graph of Fig. 7.2 with Marginalization.

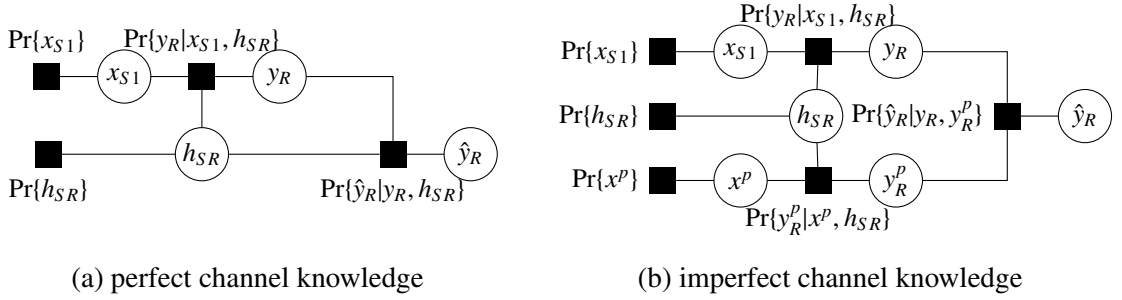


Figure 7.4: Factor graph representation of the quantizer design for one subcarrier including the compression mapping of the quantizer itself.

For the quantizer with mapping $\Pr\{\hat{y}_R|y_R, y_R^p\}$ in Fig. 7.1 for example, the pmf $\Pr\{x_{S1}, y_R, y_R^p\}$ would be required in an IB algorithm if side information for Wyner-Ziv coding is omitted. Factorizing this pmf yields

$$\begin{aligned}
 \Pr\{x_{S1}, y_R, y_R^p\} &= \sum_{h_{SR} \in \mathcal{H}_{SR}} \sum_{x^p \in \mathcal{X}^p} \Pr\{y_R|x_{S1}, h_{SR}\} \Pr\{y_R^p|x^p, h_{SR}\} \Pr\{x_{S1}\} \Pr\{h_{SR}\} \Pr\{x^p\} \\
 &= \sum_{h_{SR} \in \mathcal{H}_{SR}} \Pr\{y_R|x_{S1}, h_{SR}\} \Pr\{y_R^p|x^p, h_{SR}\} \Pr\{x_{S1}\} \Pr\{h_{SR}\}
 \end{aligned} \tag{7.3}$$

where \mathcal{X}^p has only one entry since the pilot is not actually random. Hence, the sum over all $x^p \in \mathcal{X}^p$ and $\Pr\{x^p\} = 1$ can be canceled. The pilot x^p is even so notated as random variable to visualize it in the factor graph in Fig. 7.4(b). As not only imperfect but also perfect channel knowledge is analyzed, y_R^p is substituted by h_{SR} which leads to the following pmf and its factorization.

$$\Pr\{x_{S1}, y_R, h_{SR}\} = \Pr\{y_R|x_{S1}, h_{SR}\} \Pr\{x_{S1}\} \Pr\{h_{SR}\} \tag{7.4}$$

The pmfs $\Pr\{y_R|x_{S1}, h_{SR}\}$ and $\Pr\{h_{SR}\}$ are assumed to be appropriately sampled from Gaussian processes $\mathcal{CN}(\sqrt{P_{S1}}x_{S1}, |h_{SR}|^{-2})$ and $\mathcal{CN}(0, d_{SR}^{-\alpha})$ with a resolution l_{ad} defined by the A/D converters in Fig. 7.1. The respective factor graph for both perfect and imperfect channel knowledge is depicted in Fig. 7.4 whereby the quantizer mapping $\Pr\{\hat{y}_R|y_R, h_{SR}\}$ (or $\Pr\{\hat{y}_R|y_R, y_R^p\}$) delivering \hat{y}_R is also included. Usually, factor and variable nodes could be placed as in Fig. 7.2. In

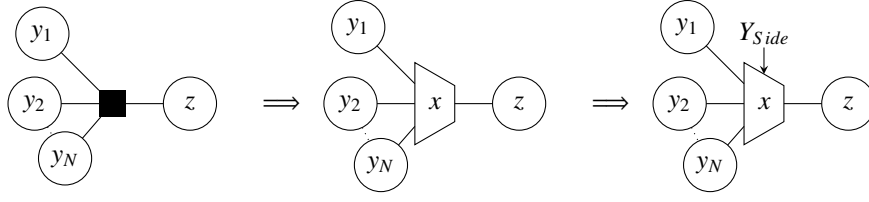


Figure 7.5: Transformation of factor graph into simple IBG and further modification to consider side information in a Wyner-Ziv coding manner.

Fig. 7.4 however, nodes are placed from a signal processing point of view with quantization as the final step. So far, these factor graphs only visualize the statistics that are required to apply the IB method. In the following, these graphs will be extended such that the compression factor is highlighted to be obtained via an IB algorithm.

7.1.2 Information Bottleneck Graphs

Originally introduced by Lewandowsky et al. in [LSB16], Information Bottleneck Graphs (IBGs) are a powerful tool to design cascaded quantizers for applications where optimal quantizer design is not feasible due to complexity issues. An IBG is simply an extension of a factor graph such that the flow of relevant information through a factor node is highlighted whereby the resulting IBG is still a factor graph.

General Functionality

Whenever a factor node represents a compression mapping $\Pr\{z|\mathbf{y}\}$ which has been optimized by the IB method (cf. Sec. 6.1.1), it is depicted as a trapezoid instead of the usually used square. Fig. 7.5 depicts how an IBG is obtained from a factor graph whereby the resulting IBG is still a factor graph itself. It simply extends the graphical presentation to emphasize the compression which can be considered as squeezing $\mathbf{y} = [y_1, \dots, y_N]$ through the bottleneck $\Pr\{z|\mathbf{y}\}$ to obtain z while keeping as much information as possible about the relevant variable x . Hence, such factor nodes are called IB nodes whereby the relevant variable is written inside the trapezoid and the compressed variable node is connected to the shortest edge of the trapezoid. All other connected variables (circles) will be compressed into z according to $\Pr\{z|\mathbf{y}\}$, i.e., multivariate compression is generally possible. Extending the IBG in the middle of Fig. 7.5 by an additional connection Y_{side} , delivers the right one. This random variable represents the side information and is therefore not declared as a variable node in the graph but simply visualizes that the respective statistics are exploited within the IB algorithm. Algorithms 1–4 can be directly applied to determine this mapping whereby the input pmf is directly given by all involved variables except the resulting compression variable, i.e., $\Pr\{x, \mathbf{y}\}$ (without side information) or $\Pr\{x, \mathbf{y}, y_{side}\}$

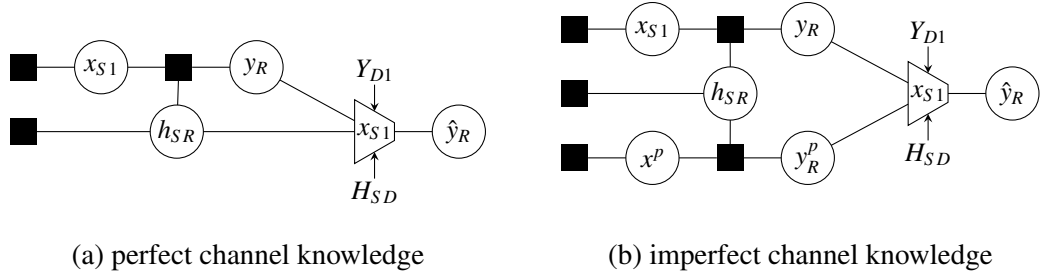


Figure 7.6: IBG representation of factor graphs in Fig. 7.4.

(with side information). For the application of these algorithms, it is important to note that multivariate compression (or as well multivariate side information) variables are stacked into a vector, e.g., $\mathbf{y} = [y_1, \dots, y_N]$.

Principal Graph for 3-Node Relay Channel

In Fig. 7.4, the statistics to apply the IB method are depicted as a factor graph which shall be now transformed into an IBG. Therefore, the factor node representing the quantizer, more precisely, its compression mapping, is replaced by an IB node (trapezoid with the relevant variable inside, cf. Fig. 7.6). From this IB node in Fig. 7.6(a) one can see directly that equalization is inherently applied because the compression mapping is optimized with respect to x_{S1} which is indicated inside the trapezoid. Furthermore, the connected side information (Y_{D1} and H_{SD}) illustrates that the optimization exploits the statistics of the destination's receive signal as well as the corresponding channel coefficient in a Wyner-Ziv source coding manner (cf. Sec. 6.1.2), i.e., the compression rate r is lowered because Y_{D1} shares mutual information with X_{S1} which increases with knowledge about H_{SD} . The IBG in Fig. 7.6(b) is very similar to Fig. 7.6(a): It is only extended by the lower branch representing pilot symbol x^p and corresponding receive signal y_R^p whereby y_R^p is now connected to the compression factor $\Pr\{\hat{y}_R|y_R, y_R^p\}$ instead of h_{SR} . As a consequence, the IB node implicitly includes channel estimation additionally to equalization. Similarly as in Fig. 7.6(a), the compression mapping in Fig. 7.6(b) exploits side information H_{SD} which is still assumed to be perfect instead of noisy for the sake of comparability (cf. Sec. 7.3). In either case (perfect or imperfect channel knowledge), direct application of the IB method is not feasible due to the overwhelming cardinality (dimensionality) of the joint pmf which is required in the IB algorithm. Considering for instance Fig. 7.6(a), $\Pr\{x_{S1}, y_R, h_{SR}, y_{D1}, h_{SD}\}$ would be of interest (all involved or connected variables except output). Due to the complex valued nature of all variables, there are totally 10 dimensions, 8 of them quasi-continuous with resolution $2^{l_{ad}}$. The cardinality 2^m of x_{S1} is directly given by modulation order m . In total, the cardinality is $2^{8 \cdot l_{ad} + m}$ which is far away from storable in today's hardware, e.g., assuming even a quite coarse resolution of $l_{ad} = 6$ and $m = 2$, $\Pr\{x_{S1}, y_R, h_{SR}, y_{D1}, h_{SD}\}$ would already have 2^{50} entries which would need 8 Petabyte of storage (RAM) in double values (cf. Table 7.1). Therefore, different

simplifying approaches are introduced and compared in the following, starting with perfect and followed by imperfect channel knowledge.

7.2 Quantizer Design for Perfect Channel Knowledge

In this section, three quantizer design approaches will be considered whereby different approximations are used: In the first approach, cascading is used to decouple and reduce the joint cardinality of the pmf that would be required in Fig. 7.6(a). In particular, one IB node is decomposed into several IB nodes with respective lower complexity. Cascading makes the compression sub-optimal because the individual optimization of several IB nodes has less degrees of freedom than the joint optimization of one IB node. Additionally, marginalization is used to exploit only a part of the side information for the sake of feasibility (cf. Sec. 7.2.4). The second approach, uses explicit equalization before quantization instead of using a cascaded quantizer which allows to independently quantize real and imaginary part. Such a decomposition is obviously optimal except for the marginalization which is still required to keep exploitation of the side information feasible (cf. Table 7.1). Anyhow, having the goal of a solely integer based signal processing in mind, the implicit approach is investigated for the sake of comparing the final performance. Finally, the explicit approach is cascaded as well to reduce cardinality further to exploit the full side information without marginalization.

7.2.1 Implicit Equalization with Partial Side Information

The originally optimal compression mapping in Fig. 7.6(a) (one IB node jointly optimized) can be cascaded according to Fig. 7.7 yielding three consecutive stages (4 IB nodes individually optimized) which are highlighted with different colors. Firstly, y_R is split into real y'_R and imaginary part y''_R using the definitions $\bar{x}' = \text{Re}(h_{SR}x_{S1}) = x'_{S1}h'_{SR} - x''_{S1}h''_{SR}$ and $\bar{x}'' = \text{Im}(h_{SR}x_{S1}) = x'_{S1}h''_{SR} + x''_{S1}h'_{SR}$ which serve as respective relevant variables. The idea of this two first stage quantizers is to preserve as much information as possible about channel coefficient and transmit signal. In this way, the intermediate compressions $y'_R \rightarrow w_1$ and $y''_R \rightarrow w_2$ lower the complexity of the second stage (cf. Table 7.1). This second stage IB node represents an inherent implicit equalization because it takes the channel coefficient together with w_1 and w_2 which are informative about both h_{SR} and x_{S1} but preserves only the relevant information about x_{S1} . The third stage, only processes the side information \bar{Y}_{D1} which is the equalized received signal at the destination. As equalization at the destination is done in any case, it does generally not matter if Y_{D1} (and H_{SD}) or \bar{Y}_{D1} (and $|H_{SD}|$) are exploited. The latter however is preferable if there is a need for marginalization: Despite the cascading into different stages, the complexity of the last stage would still be too large if the full side information is considered (cf. Sec. 7.2.4). Thus, it is

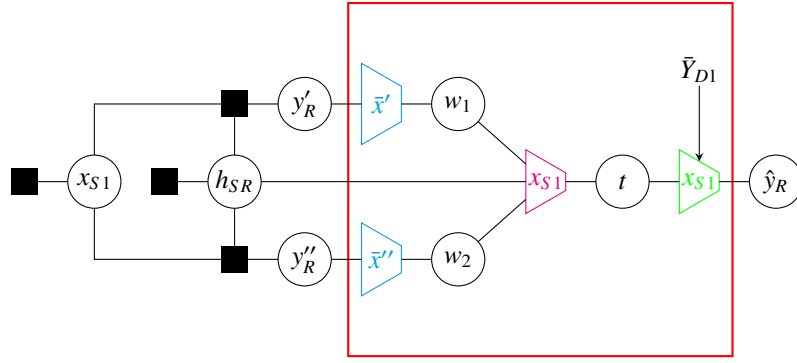


Figure 7.7: IBG with cascading into 3 stages considering part of side information.

more suitable to consider \bar{Y}_{D1} and marginalize over $|H_{SD}|$ than the other way round because \bar{Y}_{D1} shares more mutual information with X_{S1} than Y_{D1} whenever information is modulated into the phase. The respective pmfs for the IB optimization of the IB nodes are given in the following while the corresponding complexity is shown in Table 7.1.

Stage 1

If the relevant information in the first stage would be x_{S1} as usual, $y'_R = x'_{S1}h'_{SR} - x''_{S1}h''_{SR} + n'_R$ and $y''_R = x'_{S1}h''_{SR} + x''_{S1}h'_{SR} + n''_R$ could not be compressed independently since both depend on real x'_{S1} and imaginary part x''_{S1} of the relevant variable. That is why $\bar{x} = x_{S1}h_{SR}$, split into $\bar{x}' = x'_{S1}h'_{SR} - x''_{S1}h''_{SR}$ and $\bar{x}'' = x'_{S1}h''_{SR} + x''_{S1}h'_{SR}$, is defined as relevant information. As the respective real and imaginary parts, x'_{S1} and x''_{S1} , h'_{SR} and h''_{SR} as well as n'_R and n''_R , are independent and identically symmetrically distributed, the pmfs $\Pr\{\bar{x}', y'_R\}$ and $\Pr\{\bar{x}'', y''_R\}$ and, thus, quantizer mappings $\Pr\{w_1|y'_R\}$ and $\Pr\{w_2|y''_R\}$ are identical too. The required input pmf for the real part is

$$\Pr\{\bar{x}', y'_R\} = \Pr\{y'_R|\bar{x}'\} \Pr\{\bar{x}'\} \quad (7.5)$$

with $\Pr\{y'_R|\bar{x}'\}$ being appropriately sampled from a Gaussian process $\mathcal{N}\left(\sqrt{P_{S1}} \cdot \bar{x}', \frac{1}{2}\right)$ and

$$\begin{aligned} \Pr\{\bar{x}'\} &= \sum_{x_{S1}} \Pr\{\bar{x}'|x_{S1}\} \Pr\{x_{S1}\} \\ &= \sum_{x_{S1}} \left((\Pr\{h'_{SR}\} \cdot x'_{S1}) * (\Pr\{h''_{SR}\} \cdot x''_{S1}) \right) \Pr\{x_{S1}\} \end{aligned} \quad (7.6)$$

where $\Pr\{\bar{x}'|x_{S1}\}$ is obtained similarly as before from a Gaussian process $\mathcal{N}\left(0, \frac{d_{SR}^{-\alpha}}{2}|x_{S1}|^2\right)$. This becomes clear when analyzing $\bar{x}' = x'_{S1}h'_{SR} - x''_{S1}h''_{SR}$ given that x_{S1} is known: Both summands are Gaussian processes (with $\mathcal{N}\left(0, \frac{d_{SR}^{-\alpha}}{2}\right)$, cf. Sec. 2.1.2) scaled according to the given values of

x'_{S1} and x''_{S1} , that is, multiplying variances $\frac{d_{SR}^{-\alpha}}{2}$ with $|x'_{S1}|^2$ and $|x''_{S1}|^2$, respectively. Summing two Gaussians leads again to a Gaussian whereby the individual variances sum up as well.

Stage 2

For the second stage IB node,

$$\Pr\{x_{S1}, h_{SR}, w_1, w_2\} = \sum_{\mathcal{Y}'_R} \sum_{\mathcal{Y}''_R} \Pr\{w_1|y'_R\} \Pr\{w_2|y''_R\} \Pr\{y'_R|x_{S1}, h_{SR}\} \Pr\{y''_R|x_{S1}, h_{SR}\} \Pr\{h_{SR}\} \Pr\{x_{S1}\} \quad (7.7)$$

is needed where $\Pr\{y'_R|x_{S1}, h_{SR}\}$ as well as $\Pr\{y''_R|x_{S1}, h_{SR}\}$ are sampled from a Gaussian process $\mathcal{N}(h_{SR} \sqrt{P_{S1}} x_{S1}, \frac{1}{2})$. The simplification $\Pr\{w_1|y'_R\} = \Pr\{w_1|y'_R, h_{SR}, x_{S1}\}$ holds due to Markov property that the quantizer output depends only on the quantizer input (cf. Sec. 6.1).

Stage 3

To optimize the third stage with the IB algorithm,

$$\Pr\{x_{S1}, t, \bar{y}_{D1}\} = \Pr\{t|x_{S1}\} \Pr\{\bar{y}_{D1}|x_{S1}\} \Pr\{x_{S1}\} \quad (7.8)$$

is of interest, where \bar{y}_{D1} and t are independent when x_{S1} is known due to transmission over independent links. In (7.8), $\Pr\{t|x_{S1}\} = \frac{\Pr\{x_{S1}|t\} \Pr\{t\}}{\Pr\{x_{S1}\}}$ is inherently known from the IB optimization of the previous stage and $\Pr\{\bar{y}_{D1}|x_{S1}\}$ is determined via marginalization.

$$\Pr\{\bar{y}_{D1}|x_{S1}\} = \sum_{|h_{SD}| \in \mathcal{H}_{SD}} \Pr\{\bar{y}_{D1}||h_{SD}|, x_{S1}\} \Pr\{|h_{SD}|\} \quad (7.9)$$

According to equalization as in (2.13c), $\Pr\{\bar{y}_{D1}||h_{SD}|, x_{S1}\}$ is sampled from a complex Gaussian process $\mathcal{CN}(\sqrt{P_{S1}} x_{S1}, |h_{SD}|^{-2})$ and $\Pr\{|h_{SD}|\}$ from a Rayleigh process $\mathcal{R}(d_{SD}^{-\alpha})$.

7.2.2 Explicit Equalization with Partial Side Information

In Sec. 7.2.1, equalization is done implicitly within an IB node with the advantage of no additional processing except an FFT which cannot be avoided.⁴ Due to phase rotation of the channel however, I and Q component are not statistically independent and, therefore, cannot be compressed independently. Applying equalization explicitly before quantization, recovers statistical

⁴Theoretically, the goal is a vector quantizer in time domain which would have lower complexity because of sparing the FFT. Unfortunately, such an approach can neither determined nor implemented in state-of-the-art hardware because the complexity is growing exponentially with the number of subcarriers.

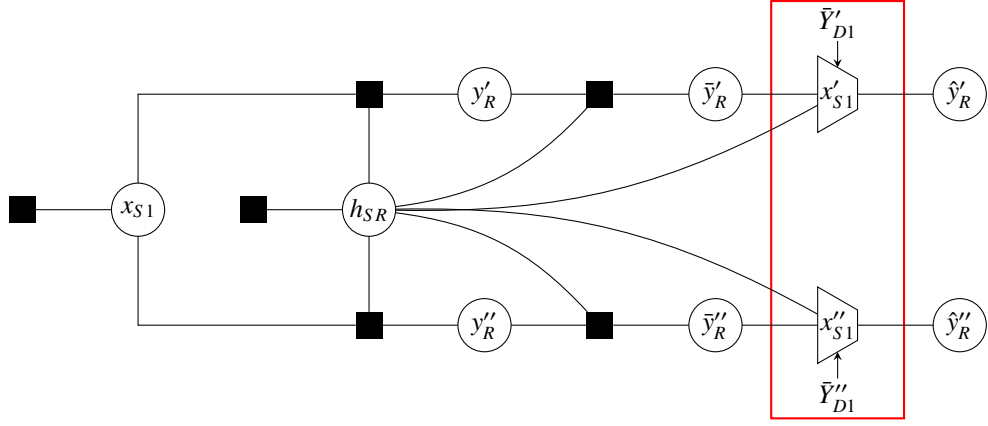


Figure 7.8: IBG with explicit equalization considering part of the side information.

independence of I and Q. Fig. 7.8 shows the modified IBG whereby the red box illustrates the splitting into real and imaginary part. Hence, this quantization approach is optimal except for the neglected side information due to marginalization of $|H_{SD}|$. More precisely, the IB algorithms have more degrees of freedom to optimize the quantizer mapping due to availability of all contributing variables in the joint input pmf. The required input pmf for the IB optimization is

$$\Pr\{x'_{S1}, \bar{y}'_R, |h_{SR}|, \bar{y}'_{D1}\} = \Pr\{\bar{y}'_R|x'_{S1}, |h_{SR}|\} \Pr\{\bar{y}'_{D1}|x'_{S1}\} \Pr\{|h_{SR}|\} \Pr\{x'_{S1}\}. \quad (7.10)$$

where $\Pr\{\bar{y}'_R|x'_{S1}, |h_{SR}|\}$ is sampled appropriately from a Gaussian $\mathcal{N}\left(\sqrt{P_{S1}}x'_{S1}, \frac{|h_{SR}|^{-2}}{2}\right)$ and $\Pr\{|h_{SR}|\}$ from a Rayleigh process $\mathcal{R}(d_{SR}^{-\alpha})$. Similar as in (7.9), $\Pr\{\bar{y}'_{D1}|x'_{S1}\}$ is obtained via marginalization with respect to $|h_{SD}|$ for the sake of feasibility.

$$\Pr\{\bar{y}'_{D1}|x'_{S1}\} = \sum_{|h_{SD}| \in \mathcal{H}_{SD}} \Pr\{\bar{y}'_{D1}||h_{SD}|, x'_{S1}\} \Pr\{|h_{SD}|\} \quad (7.11)$$

As only the real part is of interest here, $\Pr\{\bar{y}'_{D1}||h_{SD}|, x'_{S1}\}$ is sampled from $\mathcal{N}\left(\sqrt{P_{S1}}x'_{S1}, \frac{|h_{SD}|^{-2}}{2}\right)$.

7.2.3 Explicit Equalization with Full Side Information

So far the statistics of $|H_{SD}|$ have been neglected via marginalization for the sake of complexity. Furthermore, it has been shown that explicit equalization allows independent quantization of components I and Q which is feasible without cascades. Introducing cascades in branches I and Q again as depicted in Fig. 7.9, enables consideration of side information $|H_{SD}|$ additionally to \bar{Y}_{D1} . Although the cascade leads certainly to a performance degradation, the use of the full available side information may overcompensate this approximation loss. The respective input

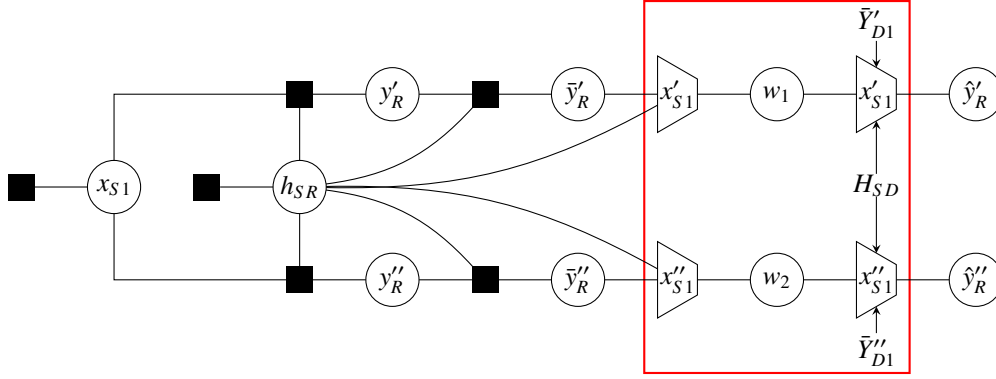


Figure 7.9: IBG with explicit equalization and cascading considering full side information.

pmfs for an IB algorithm starting with the first stage are as follows:

$$\Pr\{x'_{S1}, \bar{y}'_R, |h_{SR}|\} = \Pr\{\bar{y}'_R | x'_{S1}, |h_{SR}|\} \Pr\{|h_{SR}|\} \Pr\{x'_{S1}\} \quad (7.12a)$$

$$\Pr\{x'_{S1}, w_1, \bar{y}'_{D1}, |h_{SD}|\} = \Pr\{\bar{y}'_{D1} | |h_{SD}|, x'_{S1}\} \Pr\{|h_{SD}|\} \Pr\{x'_{S1}, w_1\} \quad (7.12b)$$

7.2.4 Complexity

Comparing the different quantizer design approaches of this section to the optimal quantizer in Fig. 7.6(a), a distinct complexity reduction in terms of cardinality becomes visible (cf. Table 7.1). The complexity depicted in Table 7.1 refers to the cardinality (and also the corresponding RAM

Table 7.1: Complexity of Quantizer Design for Perfect Channel Knowledge

	Optimal	Approach #1			Approach #2	Approach #3	
		Stage 1	Stage 2	Stage 3		Stage 1	Stage 2
l_{ad}, l_{cv}, m	$2^{8 \cdot l_{ad} + m}$	$2^{2 \cdot l_{ad}}$	$2^{2 \cdot l_{ad} + 2 \cdot l_{cv} + m}$	$2^{2 \cdot l_{ad} + l_{cv} + m}$	$2^{3 \cdot l_{ad} + \frac{m}{2}}$	$2^{2 \cdot l_{ad} + \frac{m}{2}}$	$2^{2 \cdot l_{ad} + l_{cv} + \frac{m}{2}}$
$l_{ad} = 8, l_{cv} = 4, m = 2$	$7.4 \cdot 10^{19}$	65536	67108864	4194304	33554432	131072	2097152
Main Storage	512 EB	512 KB	512 MB	32 MB	256 MB	1 MB	16 MB
$l_{ad} = 6, l_{cv} = 4, m = 2$	$1.1 \cdot 10^{15}$	4096	4194304	262144	524288	8192	131072
Main Storage	8 PB	32 KB	32 MB	2 MB	4 MB	64 KB	1 MB

for implementation in double precision) of the respective input pmf of an iterative IB algorithm where l_{ad} and l_{cv} denote the resolution of the quasi-continuous variables (A/D conversion) and the intermediate compression variables (cascading), respectively. These pmfs are functions of all random variables involved with an IB node except the output (short edge of the trapezoid). The pmf $\Pr\{x_{S1}, y_R, h_{SR}, y_{D1}, h_{SD}\}$ to optimize the mapping $\Pr\{\hat{y}_R | y_R, h_{SR}\}$ of the principal IB node in Fig. 7.6(a) for example, has a cardinality of $2^{8 \cdot l_{ad} + m}$ which arises from 4 quasi-continuous complex variables y_R , h_{SR} , y_{D1} and h_{SD} with A/D resolution l_{ad} in I and Q as well as alphabet size 2^m of the transmit signal. Obviously, such an optimal design is infeasible with state-of-the-art main storage which is in the range of Gigabytes instead of Peta- or Exabytes as

for the exemplary parametrization in Table 7.1. In contrary, all above derived suboptimal design approaches have a feasible storage need of a couple of Megabyte. The highest storage capacity of 512 or 32 MB is required by the first approach for the IB node with implicit equalization in the 2nd Stage. The storage complexity is mainly determined by the A/D resolution l_{ad} which is with two different parametrizations $l_{ad} = 8$ and $l_{ad} = 6$ Bit shown in Table 7.1. The resolution l_{cv} of intermediate compression variables is set to $l_{cv} = 4$ Bit which is beyond the intended range of rates at the quantizer output. Although the physical limit of state-of-the-art main storage has room for a couple of Gigabytes, the A/D resolution l_{ad} is not further increased because the run time of an iterative IB algorithm would grow too large to obtain different quantizers (range of SNRs and compression rates r) in a meaningful time. For the simulation results in Sec. 7.4, a resolution of $l_{ad} = 8$ Bit is usually used except for the very complex IB node in the second stage of the first approach. Finally, as the complexity is determined by the largest pmf of an approach, the last design has not only the lowest complexity but also the lowest expectable performance degradation (if the exploitation of the full side information overcompensates the decomposition into cascades cf. Sec. 7.4).

7.3 Quantizer Design for Imperfect Channel Knowledge

Here, the approaches of the previous section will be extended such that the IB optimization takes channel estimation errors into account. For the sake of clarity, the focus will be set to the general influence of estimation errors as well as to the difference between implicit and explicit channel estimation/equalization. Therefore, the following subsections extend Sec. 7.2.1 and Sec. 7.2.2. Please note that h_{SD} is assumed to be still perfectly known to ensure a fair comparison when using a quantizer designed for perfect channel knowledge. In particular, Wyner-Ziv coding is only theoretically considered via the achievable source coding rate r but not actually simulated: Side information decreases due to noise in the estimated channel coefficient. This would lead to a benefit when pragmatically using the quantizer, optimized for perfect channel knowledge, because the simulation environment neglects errors occurring due to mismatched Wyner-Ziv coding.

7.3.1 Implicit Channel Estimation and Equalization

The implicit approach as depicted in Fig. 7.10 corresponds to the one in Fig. 7.7. The first stage is simply extended by two IB nodes compressing the received pilot signal y_R^p to intermediate compression variables w_{p1} and w_{p2} that are informative about real h'_{SR} and imaginary part h''_{SR} of the channel coefficient, respectively. These two IB nodes realize the implicit channel estimation, the second stage realizes as before the implicit equalization, and the third stage incorporates the

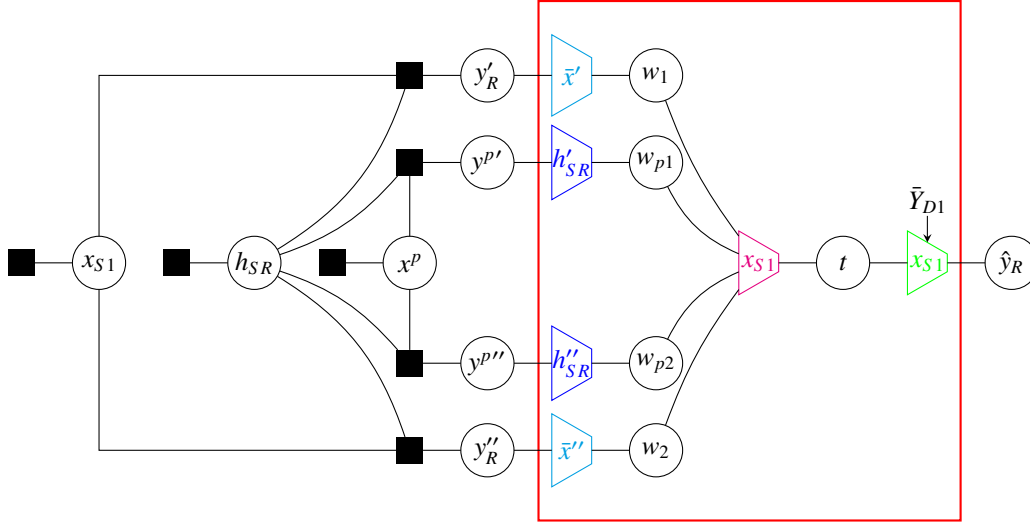


Figure 7.10: IBG with cascading into 3 stages with implicit channel estimation and equalization.

side information according to Wyner-Ziv coding. It is important to note that due to inherent compression of channel knowledge, the complexity is reduced ($2^{2 \cdot l_{ad}} \cdot 2^{2 \cdot l_{cv}} \cdot 2^m \rightarrow 2^{4 \cdot l_{cv}} \cdot 2^m$, cf. Table 7.2) in the second stage of Fig. 7.10 compared to the IBG in Fig. 7.7 which has perfect channel knowledge. The outer IB nodes in the first stage (cyan) and the IB node in the third stage (green) are identical to the IB nodes (same color) in Fig. 7.7.⁵ The pmfs to determine the new (blue) or modified (magenta) IB nodes are given in the following.

Inner IB Nodes (blue) of Stage 1

For the inner IB nodes in the first stage, due to symmetry of real and imaginary part, i.e., $\Pr\{h'_{SR}, y_R^{p'}\} = \Pr\{h''_{SR}, y_R^{p''}\}$, the required pmf is

$$\Pr\{h'_{SR}, y_R^{p''}\} = \Pr\{y_R^{p''} | h'_{SR}\} \Pr\{h'_{SR}\} \quad (7.13)$$

where $\Pr\{y_R^{p''} | h'_{SR}\}$ and $\Pr\{h'_{SR}\}$ are sampled from Gaussian processes $\mathcal{N}\left(\sqrt{P_{S1}} h'_{SR}, \frac{1}{2}\right)$ and $\mathcal{N}\left(0, \frac{d_{SR}^{-\alpha}}{2}\right)$.

⁵Reuse of the third stage IB node implies perfect knowledge of h_{SD} which is assumed to ensure a fair comparison because actual Wyner-Ziv coding is not implemented.

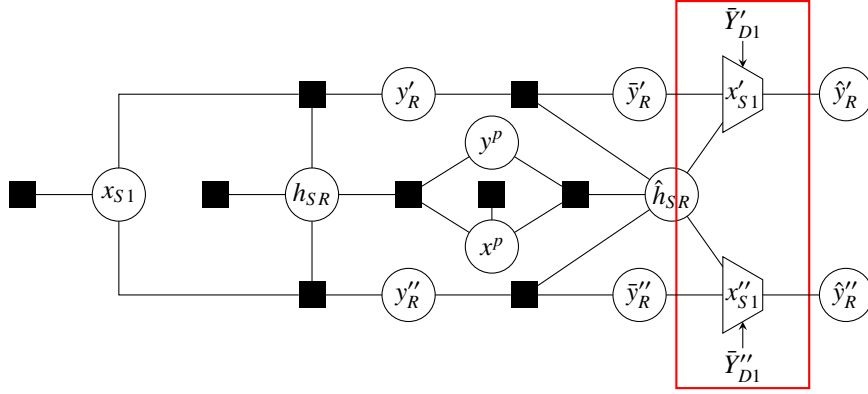


Figure 7.11: IBG with explicit channel estimation and equalization.

Stage 2 (magenta)

In the second stage, the IB algorithm needs

$$\begin{aligned} \Pr\{x_{S1}, w_1, w_2, w_{p1}, w_{p2}\} &= \sum_{y'_R \in \mathcal{Y}'_R} \sum_{y''_R \in \mathcal{Y}''_R} \Pr\{w_1|y'_R\} \Pr\{w_2|y''_R\} \\ &\quad \sum_{h_{SR} \in \mathcal{H}_{SR}} \Pr\{y'_R|h_{SR}, x_{S1}\} \Pr\{y''_R|h_{SR}, x_{S1}\} \Pr\{h_{SR}, w_{p1}, w_{p2}\} \Pr\{x_{S1}\} \end{aligned} \quad (7.14)$$

where $\Pr\{h_{SR}, w_{p1}, w_{p2}\} = \Pr\{h'_{SR}|w_{p1}\} \Pr\{h''_{SR}|w_{p2}\} \Pr\{w_{p1}\} \Pr\{w_{p2}\}$ and the mappings $\Pr\{w_1|y'_R\}$ as well as $\Pr\{w_2|y''_R\}$ are inherently known from the previous stage.

7.3.2 Explicit Channel Estimation and Equalization

The explicit approach as depicted in Fig. 7.11 corresponds to the IBG in Fig. 7.8. The only difference in the IB nodes is the consideration of \hat{h}_{SR} instead of h_{SR} . The pmf to determine the compression mapping which holds for both the upper and the lower IB node, is

$$\Pr\{x'_{S1}, \bar{y}'_R, |\hat{h}_{SR}|, \bar{y}'_{D1}\} = \Pr\{\bar{y}'_R||\hat{h}_{SR}|, x'_{S1}\} \Pr\{|\hat{h}_{SR}|\} \Pr\{x'_{S1}\} \sum_{|h_{SD}| \in \mathcal{H}_{SD}} \Pr\{\bar{y}'_{D1}||h_{SD}|, x'_{S1}\} \Pr\{|h_{SD}|\} \quad (7.15)$$

where, according to (2.13a) and (2.13b), $\Pr\{\bar{y}'_R||\hat{h}_{SR}|, x'_{S1}\}$, $\Pr\{\bar{y}'_{D1}||h_{SD}|, x'_{S1}\}$, $\Pr\{|\hat{h}_{SR}|\}$, and $\Pr\{|h_{SD}|\}$ are appropriately sampled from $\mathcal{N}(\sqrt{P_{S1}}x'_{S1}, |\hat{h}_{SR}|^{-2})$, $\mathcal{N}(\sqrt{P_{S1}}x'_{S1}, |h_{SD}|^{-2})$, $\mathcal{R}(d_{SR}^{-\alpha} + 1)$, and $\mathcal{R}(d_{SD}^{-\alpha} + 1)$, respectively

7.3.3 Complexity

The complexity of the just derived two approaches does not differ from the analysis in Sec. 7.2.4 except for the first approach with implicit channel estimation and equalization: Firstly, the IB node optimization in the second stage has lower complexity due to beforehand compression of channel knowledge. Secondly, the IB nodes (inner), which compress the observed pilots $y_R^{p'}$ and $y_R^{p''}$ with respect to the channel coefficients h'_{SR} and h''_{SR} , are added whereby their complexity does not differ from that of the outer IB nodes.

Table 7.2: Complexity of Quantizer Design for Imperfect Channel Knowledge

	Optimal	Approach #1				Approach #2
		Stage 1 (outer)	Stage 1 (inner)	Stage 2	Stage 3	
l_{ad}, l_{cv}, m	$2^{8 \cdot l_{ad} + m}$	$2^{2 \cdot l_{ad}}$	$2^{2 \cdot l_{ad}}$	$2^{4 \cdot l_{cv} + m}$	$2^{2 \cdot l_{ad} + l_{cv} + m}$	$2^{3 \cdot l_{ad} + \frac{m}{2}}$
$l_{ad} = 8, l_{cv} = 4, m = 2$	$7.4 \cdot 10^{19}$	65536	65536	262144	4194304	33554432
Main Storage	512 EB	512 KB	512 KB	2 MB	32 MB	256 MB
$l_{ad} = 6, l_{cv} = 4, m = 2$	$1.1 \cdot 10^{15}$	4096	4096	262144	262144	524288
Main Storage	8 PB	32 KB	32 KB	2 MB	2 MB	4 MB

7.4 Achievable Rates for Compress and Forward Using Multiple Carriers

The rate allocation is done very similar as in Sec. 6.2.2 using (6.9) to determine the final rate for CF. The partial rates R_b^{xR} and R_b^{xS2} are as well obtained as in Sec. 2.4 with the only difference of considering effective SNRs which are obtained via averaging the mutual informations of all subcarriers. More precisely, the individual carrier-SNRs are converted into a mutual information and the average mutual information is then converted back into an effective SNR. The broadcast rate R_b^{xS1} is obtained with an extra simulation that contains the quantizers obtained with respect to the different approaches. This simulation has the same functionality as the one shown in Fig. 6.6 but is extended to an OFDM system (cf. Sec. 2.5) with $N_{FFT} = 1024$ subcarriers and a guard interval of length $N_{GI} = 256$. Again, a line setup with equal transmit power at all nodes is assumed. The SNR is always that of the direct link either with respect to individual transmit power per node or sum transmit energy.

7.4.1 Perfect Channel Knowledge

Different Quantizer Design Approaches

To investigate the performance of the different quantizer design approaches, Fig. 7.12 shows the total achievable rate R_{CF} versus SNR (power per node) on the direct link for orthogonal MAC. Despite the three beforehand derived approaches which design a quantizer for the ensemble of

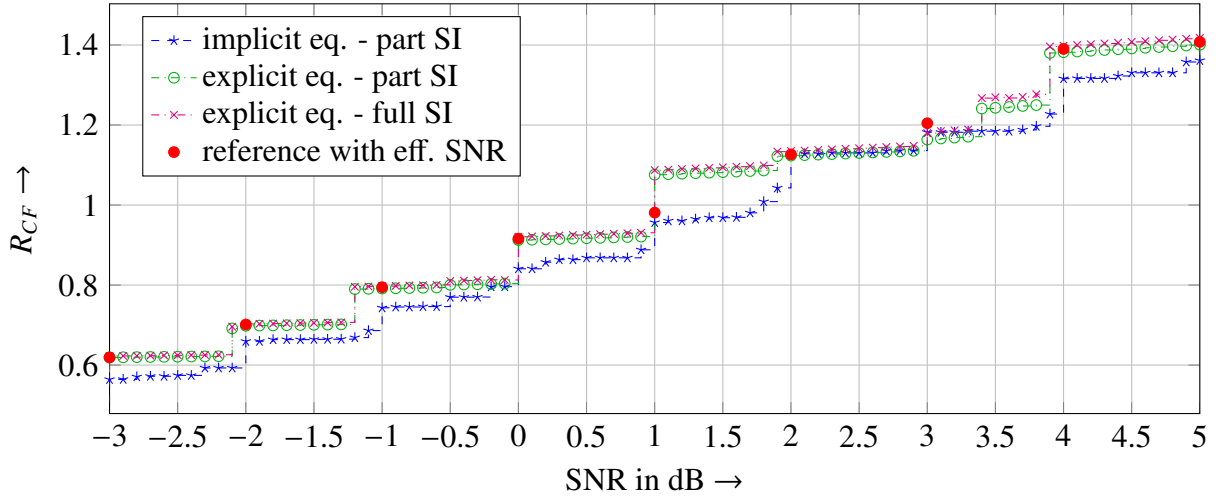


Figure 7.12: Total achievable rates for CF versus SNR (individual power) on direct link at $d = 0.8$ with $\alpha = 4$ comparing different quantizer approaches for orthogonal channel access.

channel coefficients h_{SR} , Fig. 7.12 shows additionally results (red filled circles) for the approach described in App. H which is used as a theoretical reference scheme. The SNR resolution is coarser due to the much higher effort to simulate each point in detail. Anyhow, at the comparable points, the best of derived quantizer designs has more or less equal performance as the reference scheme. Comparing the derived quantizer designs among themselves, the first approach (cf. Fig. 7.7) with implicit equalization performs obviously worst except for $\text{SNR} \in [2, 3.4]$.⁶ This bad performance is reasoned in the cascading which splits the original IB node with its full degrees of freedom into three less complex IB problems with an also less efficient compression due to reduced degrees of freedom during the optimization. Hence, explicit equalization is a key issue and provides the smallest performance degradation. The reason is simply the dimensionality reduction (without approximation) due to independent quantization of I and Q components (cf. Fig. 7.8). Furthermore, independent quantization of I and Q enables exploitation of the full side information by reintroducing cascading in the I and Q branches (cf. Fig. 7.9). The results in Fig. 7.12 reveal that consideration of the full side information compensates the suboptimal quantization due to cascades in branches I and Q. However, the gain in terms of performance is

⁶The reason for this exceptions is explained in the end of Sec. 7.4.2.

very small compared to the improvement obtained by explicit equalization. As a last point, one may think that a further joint compression of the independently quantized components I and Q could lead to an improvement since both quantizers use the channel magnitude $|h_{SR}|$ as input. This is not the case because the IB optimization focuses on preserving relevant information about x_{S1} and does not care about the actual channel amplitude. In other words, there are different combinations of $|h_{SR}|$ and y_R yielding a specific compression index \hat{y}_R subject to a specific relevant information. Practically, relevant information is expressed by probability $\Pr\{x_{S1}|\hat{y}_R\}$ which does neither directly depend on nor differentiate between y_R and $|h_{SR}|$, e.g., whether the channel is strong and the observation very noisy or the other way around.

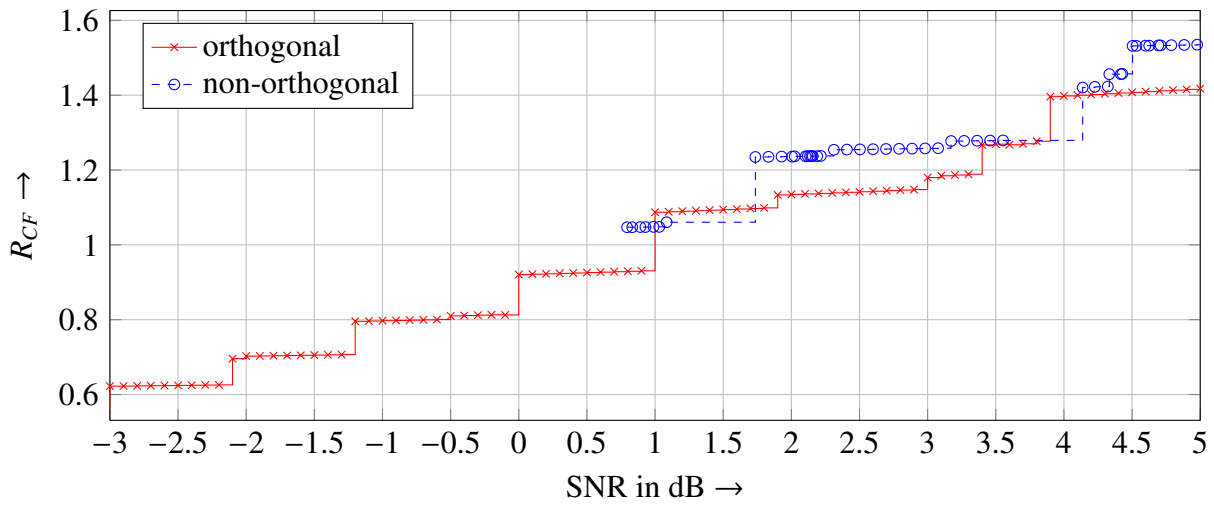


Figure 7.13: Total achievable rates for CF versus SNR (sum energy) on direct link at $d = 0.8$ with $\alpha = 4$ comparing non-orthogonal and orthogonal channel access for exploitation of total side information.

Non-orthogonal versus Orthogonal Channel Access

Fig. 7.13 shows the achievable CF rates for the quantizer design which exploits the full side information. Similar as in the last chapter, the non-orthogonal scheme does not pay off compared to the orthogonal in the low SNR range but does outperform it usually for moderate and high SNR. The reason is again the SNR on the direct link which is too low to allocate one of the available MCSs as depicted in Fig. 2.22. In some cases, the non-orthogonal scheme does not outperform the orthogonal due to the higher sum energy consumption which is reasoned in the discreteness of transmission and source coding rates.

7.4.2 Imperfect Channel Knowledge

To discuss the influence of imperfect channel knowledge on the achievable rate R_{CF} , Fig. 7.14 shows the results for quantization as in Fig. 7.8 (optimized for perfect channel knowledge) for both perfect (blue circles) and imperfect channel knowledge (blue dashed squares). Furthermore, the adapted quantizer approaches with either implicit (cf. Fig. 7.10) or explicit (cf. Fig. 7.11) estimation/equalization are shown in green (dashdotted pluses) and red (dashdotdotted crosses), respectively. The main observation in Fig. 7.14 is the distinct rate loss caused by the channel estimation errors which is present for both the mismatched and the adapted quantizer. Hence, the visible gain due to this adaptation is also very limited. The reason is simply that channel estimation errors in either case reduce the mutual information between \hat{y}_R and x_{S1} : In the mismatched case, the error occurs at the quantizer input because $|h_{SR}| \neq |\hat{h}_{SR}|$ which leads to an error in $\Pr\{x_{S1}|\hat{y}_R\}$ at the destination decoder. In the adapted case, the quantizer is matched but less efficient due to the additional noise in $|\hat{h}_{SR}|$ which decreases the ratio of preserved information to source coding rate. Thus, the forwarding of \hat{y}_R to the destination requires more time or provides less information. Either way the totally achievable rate R_{CF} is degraded. In the low SNR range, the difference between these two different errors is visible in the achievable rate, that is, a mismatched quantizer is less suited than the adapted which is expectable. Furthermore, as expected, the adapted quantizer design with explicit estimation and equalization performs best in this SNR regime. For higher SNRs however, the performance of all approaches is very similar except for the approach with cascades (dashed green pluses). Hence, even a mismatched quantizer scheme performs well and sometimes slightly better than the expected optimal design (dashed red crosses). These small variations in the achievable rate are reasoned in the artificially created discrete range of source coding rates r which limit the degrees of freedom in (6.9).⁷ More precisely, due to the restriction to deterministic quantizers, r depends only on the cardinality $|\hat{\mathcal{Y}}_R|$ of \hat{y}_R and the specific quantizer design according to an IBG. Hence, all quantizer design approaches have different sets of available rates r according to either $r = H(\hat{Y}_R|\bar{Y}_{D1}), \forall |\hat{\mathcal{Y}}_R| \in \{2, \dots, 16\}$ for complex compression or $r = H(\hat{Y}'_R|\bar{Y}'_{D1}) + H(\hat{Y}''_R|\bar{Y}''_{D1}), \forall |\hat{\mathcal{Y}}'_R| = |\hat{\mathcal{Y}}''_R| \in \{2, \dots, 16\}$ for independent compression of I and Q. These available rates r are shown in the last plot of Fig. 7.14 for $|\hat{\mathcal{Y}}'_R| = |\hat{\mathcal{Y}}''_R| \in \{2, \dots, 5\}$ and reveal that the slightly better performance of the unadapted quantizer is only caused due to the possibility to pick a slightly smaller r (3rd plot) which allows a slightly higher τ (4th plot). Please note that this effect can only occur when a jump to a better transmission rate $R_b^{x_{S1}}$ (2nd plot), which are constant in a specific SNR range, is not possible. The same holds for the green curve which is not plotted in all subplots for the sake of clarity.

⁷The same is true for the transmission rates that are determined by the MCSs. Anyhow, this is not relevant here because all quantizer approaches are simulated with the same BICM setup.

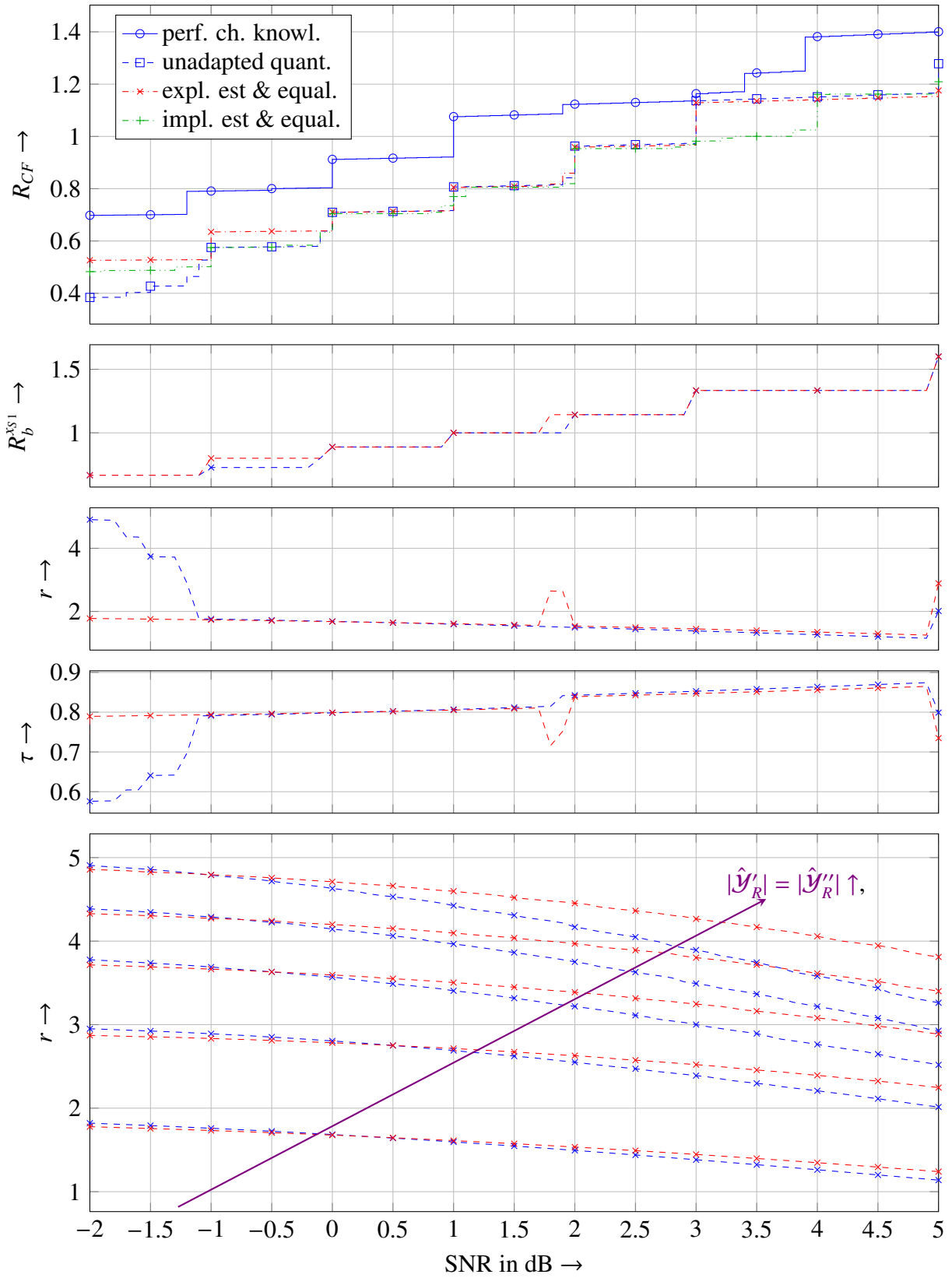


Figure 7.14: Total achievable rates for CF versus SNR (individual power) on direct link at $d = 0.8$ with $\alpha = 4$ comparing different quantizer approaches for orthogonal channel access.

7.5 Discussion

In contrary to all previous chapters which consider single carrier transmission according to an AWGN channel model with an attenuation coefficient due to path-loss, this chapter has considered multi-carrier transmission with Rayleigh fading subcarriers to investigate the influence of randomly changing channel coefficients on the quantizer design. There are two possibilities to exploit the channel knowledge during quantization: On one hand, equalization may be applied before quantization whereby the channel influence is present in the SNR per carrier for which the quantizer is designed as in the previous chapter. On the other hand, the channel coefficient may be considered as additional input, that is, quantization is designed for the ensemble of possible channel coefficients. Both approaches have their drawbacks and benefits and are generally infeasible. The first case is not implementable using deterministic quantization due to dependence of the compression rate on the instantaneous vector channel realization. Random quantization could avoid this issue but is not feasible due to an additionally required bisection search. Either way deterministic quantization is preferable since it can be implemented straightforward as LUT. Hence, quantizers for the ensemble of channel coefficients have been derived while the other approach has been used as reference. As quantizer design for the ensemble of channel coefficients is generally too complex due to storage incapability of the required pmfs, IBGs have been used to find suboptimal but feasible quantizer designs. A couple of approaches with differing approximations were derived for perfect and imperfect channel knowledge and compared in terms of simulation results. According to these results, explicit equalization should be applied before quantization to achieve a more efficient compression and, thus, a better total performance. Furthermore, it could be shown that exploitation of the full side information instead of marginalizing the channel amplitude at the cost of cascading is preferable and performs equally good as the reference scheme. Finally, imperfect channel knowledge degrades the performance as expected. However, adapting the quantizer design to the statistics of the channel estimation errors yields only a significant improvement for low SNRs. In the high SNR regime, unadapted quantizers perform very similar.

Chapter 8

Conclusion

Two very common relay protocols, Decode-and-Forward (DF) and Compress-and-Forward (CF) [CE79, KGG05, HMZ05], have been studied from a practical point of view: On one hand, theoretically neglected imperfections due to practical implementation like estimation errors have been considered. On the other hand, the question of how to implement specific signal processing steps has been addressed. In particular, the actual quantizer design for CF, which is only modeled as quantization noise in the information theoretic analysis in Ch. 3, has been accomplished and evaluated. Despite some information theoretic results in Ch. 3, Ch. 4 and Ch. 5, simulation results have been determined for a Bit Interleaved Coded Modulation (BICM) system using state-of-the-art coding and modulation techniques (turbo coding, Quadrature Amplitude Modulation (QAM)). As a practical matter, half-duplex relays have been considered which enables to differentiate between general non-orthogonal and orthogonal channel access in the second time slot: The relay may transmit either alone (orthogonal access) or jointly with the source (non-orthogonal access) which requires more complicated signal processing. Both channel access schemes have been compared throughout the work with respect to individual transmit power per node as well as sum sum energy.

The main part of Ch. 3 reviews information theoretic results from the literature [KGG05, HMZ05, Wei12] which reveal a huge gain of non-orthogonal channel access towards orthogonal access for DF whenever distributed beamforming is reasonable, that is, source-destination and relay-destination link have comparable quality. If beamforming is not meaningful, this gain is still moderate for an individual power constraint as it is usually considered in literature. At the end of Ch. 3 however, it has been shown that, for an individual and a sum energy constraint, this gain is only visible for distributed beamforming but vanishes otherwise [KK15a]. Motivated by transmission strategy of the source in the Multiple Access (MAC) phase to either apply distributed beamforming, superposition coding, or a mixture of both, Ch. 4 has generalized this information theoretic concept into a multilevel code. Due to the superposition nature of the relay channel, superposition modulation [Wo11] was used to create the desired mapping.

The usually two bit levels have been furthermore extended to arbitrary many layers to exploit a shaping gain [DRU97] which allows to approach the information theoretically achievable rate. Moreover, bit level design was adapted such that the resulting joint mapping of two bit streams at the destination follows a specific power allocation scheme presented in Sec. 2.3.4. As a consequence, the destination can demap both bit streams at once with low complexity instead of applying successive interference cancellation. This adapted superposition multilevel code performs almost as well as without the restriction but cannot be applied for each imaginable scenario [KK15b].

The next topic which has been discussed, is imperfect phase knowledge which is necessary to apply distributed beamforming. More precisely, the source must know the channel phase of the relay's transmit signal to ensure a constructive superposition at the destination. In a practical usually noisy system, phase estimation errors will occur and degrade the performance. The arising question is how strong this impact is and up to which point beamforming is still reasonable. Until now, phase errors have been modeled either as negligible or uniformly distributed which disables beamforming [KGG05]. In Ch. 5, a better suited model for phase errors based on a Phase Locked Loop (PLL) has been introduced and used to investigate the impact of such errors on the achievable rate for both information theoretic and simulation results: As expectable, phase errors degrade the performance in dependence of the Signal to Noise Ratio (SNR) of a first order PLL. If this SNR is low, severe phase errors cause disabling of distributed beamforming while for high SNR, distributed beamforming still pays off with only minor degradation. Simulation results have mainly confirmed these insights [KK15c]. However, due to the limited range of Modulation and Coding Scheme (MCS) especially in the low SNR regime, orthogonal channel access may be preferable when there is no suitable MCS to transmit an additional superposed signal from the source, i.e., when the rate of the smallest MCS already exceeds the virtual capacity.

The last two chapters have dealt with quantizer design for CF whereby Ch. 6 has considered path loss attenuated Additive White Gaussian Noise (AWGN) links and Ch. 7 frequency selective Rayleigh fading channels decomposed into multiple carriers via Orthogonal Frequency Division Multiplexing (OFDM). In Ch. 6, the Information Bottleneck (IB) method [TPB99, Slo02, ZKBW09a, ZBW12, Zei12] has been used to design optimal quantizers for the given setup which are implemented as Look-Up-Tables (LUTs) in a BICM system. The simulation results confirm functionality and a well performance especially when compared to the same results for DF [KK16]: CF is indeed superior when the source-relay link is the bottleneck of the system. Insights with respect to channel access are similar as for DF. In Ch. 7, the plain IB method cannot be directly applied as in Ch. 6 because the implementation would require far too much storage. Thus, Information Bottleneck Graphs (IBGs) [LSB16] have been introduced to make quantizer design feasible. Different approximations were used to derive three quantizer

designs whose performance has been compared again via simulation results [KK17a, KK17c]. It could be shown that all approaches work and that explicit equalization before quantization yields the best performance comparable to the reference scheme. Finally, imperfect channel knowledge has been investigated [KK17b]: Results reveal that channel estimation error statistics should be considered in the quantizer design, at least for high noise variances. Again, explicit channel estimation and equalization yields better results.

Bibliography

- [ACLY00] Ahlswede, Rudolf, Cai, Ning, Li, Shuo Yen Robert, and Yeung, Raymond W.: *Network Information Flow*. IEEE Transactions on Information Theory, 46(4):1204–1216, 2000. 1
- [BCJR74] Bahl, L., Cocke, J., Jelinek, F., and Raviv, J: *Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate*. IEEE Transactions on Information Theory, (January 1972):284–287, 1974. 11, 143
- [BGT93] Berrou, Claude, Glavieux, Alain, and Thitimajshima, Punya: *Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes (1)*. (1), 1993. 9, 10
- [Bla72] Blahut, Richard E.: *Computation of channel capacity and rate-distortion functions*. IEEE Transactions on Information Theory, 18(4):460–473, 1972, ISSN 0018-9448. 70
- [BSM05] Bronstein, Ilja N, Semendjajew, Konstantin A, and Musiol, Gerhard: *Taschenbuch der Mathematik*. Verlag Harri Deutsch, 6., vollst edition, 2005, ISBN 3-817-12006-0. 62
- [CE79] Cover, Thomas and El Gamal, Abbas: *Capacity Theorems for the Relay Channel*. IEEE Transactions on Information Theory, IT-25(5):572–584, 1979. 1, 27, 28, 30, 31, 32, 34, 111
- [CT91] Cover, Thomas M and Thomas, Joy A: *Elements of Information Theory*. Wiley-Interscience, New York, NY, USA, 1991, ISBN 0-471-06259-6. 1, 7, 12, 25, 27, 28, 35, 72, 121
- [DRU97] Duan, Long, Rimoldi, B., and Urbanke, R.: *Approaching the AWGN Channel Capacity without Active Shaping*. Proceedings of IEEE International Symposium on Information Theory, (2):374, 1997. 14, 112
- [DSZ⁺11] Dietl, Guido, Sciora, Matthieu, Zeitler, Georg, Bauch, Gerhard, and Widmer, Joerg: *A Quantize-and-Forward Scheme for Future Wireless Relay Networks*. pages 8–11, 2011. 90

- [EK12] El Gamal, Abbas and Kim, Young Han: *Network Information Theory*. Cambridge University Press, 2012, ISBN 1107008735. 1, 25, 27, 28, 30, 31, 32, 33, 34, 35, 125
- [FF56] Ford, LR and Fulkerson, DR: *Maximal flow through a network*. Canadian Journal of Mathematics, 1956. 1, 25, 27, 28
- [FF62] Ford, L R and Fulkerson, D R: *Flows in Networks*. Princeton University Press, 1962. 1, 25, 27, 28
- [Gal63] Gallager, Robert G: *Low Density Parity Check Codes*. MIT Press, Cambridge, 1963. 9
- [HM02] Host-Madsen, Anders: *On the Capacity of Wireless Relaying*. Proceedings IEEE 56th Vehicular Technology Conference, 3(5):1333–1337, 2002. 1, 27
- [HMZ05] Host-Madsen, Anders and Zhang, Junshan: *Capacity Bounds and Power Allocation for Wireless Relay Channels*. IEEE Transactions on Information Theory, 51(6):2020–2040, 2005. 1, 27, 29, 35, 36, 111
- [HW11] Höher, Peter Adam and Wo, Tianbin: *Superposition Modulation: Myths and Facts*. IEEE Communications Magazine, 49(December):110–116, 2011. 17, 18, 21
- [IH77] Imai, H. and Hirakawa, S.: *A New Multilevel Coding Method Using Error-Correcting Codes*. Information Theory, IEEE Transactions on, 23(3):371–377, may 1977, ISSN 0018-9448. 12
- [JBS00] Jeruchim, Michel C, Balaban, Philip, and Shanmugan, K Sam: *Simulation of Communication Systems: Modeling, Methodology and Techniques*. Kluwer Academic Publishers, Norwell, MA, USA, 2nd edition, 2000, ISBN 0306462672. 56, 57
- [Kam08] Kammeyer, Karl Dirk: *Nachrichtenübertragung*, chapter 16.6, pages 628–635. Vieweg+Teubner Verlag, 4. edition, 2008, ISBN 978-3-8351-0179-1. 22
- [KFL01] Kschischang, Frank R., Frey, Brendan J., and Loeliger, Hans Andrea a.: *Factor Graphs and Sum Product Algorithm*. IEEE Transactions on Information Theory, 47(2):498–519, 2001, ISSN 00189448. 92
- [KGG05] Kramer, G., Gastpar, M., and Gupta, P.: *Cooperative Strategies and Capacity Theorems for Relay Networks*. IEEE Transactions on Information Theory, 51(9):3037–3063, sep 2005, ISSN 0018-9448. 1, 2, 27, 28, 30, 31, 32, 34, 35, 38, 111, 112
- [KK15a] Kern, Daniel and Kühn, Volker: *Achievable Decode and Forward Rates for the 3-Node Relay Channel Considering Different Energy Constraints and Synchro-*

- nization Errors. In *12th International Symposium on Wireless Communication Systems (ISWCS'15)*, pages 1–5, Brussels, Belgium, aug 2015. 27, 40, 111
- [KK15b] Kern, Daniel and Kühn, Volker: *On Superposition Multilevel Coding for the 3-Node Relay Channel Exploiting Decode and Forward*. In *IEEE 82nd Vehicular Technology Conference (VTC 2015 Fall)*, Boston, USA, sep 2015. 2, 47, 112
- [KK15c] Kern, Daniel and Kühn, Volker: *Practical Aspects of Decode and Forward in the Classical Relay Channel Exploiting Non-Orthogonal Channel Access*. International ITG Conference on Systems, Communications and Coding, 2015. 2, 55, 112
- [KK16] Kern, D and Kühn, V: *Practical Aspects of Compress and Forward with BICM in the 3-Node Relay Channel*. In *20th International ITG Workshop on Smart Antennas (WSA 2016)*, Munich, Germany, mar 2016. 69, 82, 112
- [KK17a] Kern, D and Kühn, V: *On Compress and Forward with Multiple Carriers in the 3-Node Relay Channel Exploiting Information Bottleneck Graphs*. In *11th International ITG Conference on Systems, Communications and Coding (SCC 2017)*, Hamburg, Germany, 2017. 2, 90, 113
- [KK17b] Kern, D and Kühn, V: *On Implicit and Explicit Channel Estimation for Compress and Forward Relaying OFDM Schemes Designed by Information Bottleneck Graphs*. In *21st International ITG Workshop on Smart Antennas (WSA 2017)*, Berlin, Germany, 2017. 2, 90, 113
- [KK17c] Kern, Daniel and Kühn, Volker: *On Information Bottleneck Graphs to Design Compress and Forward Quantizers Exploiting Side Information for Multi-Carrier Transmission*. In *IEEE International Conference on Communications*, Paris, France, 2017. 2, 90, 113
- [Ksc03] Kschischang, Frank R.: *Codes defined on graphs*. IEEE Communications Magazine, 41(8):118–125, 2003, ISSN 01636804. 92
- [Kur17] Kurkoski, Brian M.: *On the Relationship Between the KL Means Algorithm and the Information Bottleneck Method*. 11th International ITG Conference on Systems, Communications and Coding (SCC2017, 2017. 73
- [Lan02] Laneman, J. Nicholas: *Cooperative diversity in wireless networks: algorithms and architectures*. (1995), 2002. 1, 30, 36
- [LB15] Lewandowsky, Jan and Bauch, Gerhard: *Trellis based node operations for LDPC decoders from the information bottleneck method*. 2015. 89
- [LCG⁺09] Li, Jiaxiang, Chen, Qingchun, Gao, Suyue, Ma, Zheng, and Fan, Pingzhi: *The Optimal Puncturing Pattern Design for Rate-Compatible Punctured Turbo Codes*.

- Wireless Communications & Signal Processing, (104026):6–10, 2009. 10
- [LSB16] Lewandowsky, Jan, Stark, Maximilian, and Bauch, Gerhard: *Information Bottleneck Graphs for Receiver Design*. IEEE International Symposium on Information Theory (ISIT), 2016. 2, 75, 89, 90, 92, 94, 112
- [LTW04] Laneman, J. Nicholas, Tse, David N. C., and Wornell, Gregory W.: *Cooperative diversity in wireless networks: Efficient protocols and outage behavior*. IEEE Transactions on Information Theory, 50(12):3062–3080, 2004. 1, 30, 36
- [MBM07] Mudumbai, R, Barriac, G., and Madhow, U.: *On the Feasibility of Distributed Beamforming in Wireless Networks*. IEEE Transactions on Wireless Communications, 6(5):1754–1763, 2007. 1
- [MP04] Ma, X. and Ping, L.: *Coded Modulation Using Superimposed Binary Codes*. IEEE Transactions on Information Theory, 50(12):3331–3343, dec 2004, ISSN 0018-9448. 15
- [NFK07] Neuber, A, Freudenberger, J, and Kühn, Volker: *Coding Theory - Algorithms, Architectures and Applications*. John Wiley & Sons Ltd., UK, oct 2007. 9, 10
- [Sha48] Shannon, C. E.: *A Mathematical Theory of Communication*. The Bell System Technical Journal, 27(July):379–423, 1948. 7, 25
- [Sho03] Shokrollahi, Amin: *LDPC Codes: An Introduction*. 2003. 9
- [Slo02] Slonim, Noam: *The information bottleneck: Theory and applications*. Dissertation, 2002. 2, 69, 70, 112
- [Sti09] Stierstorfer, Clemens: *A Bit-Level-Based Approach to Coded Multicarrier Transmission*. PhD thesis, 2009. 21
- [SW73] Slepian, David and Wolf, Jack K.: *Noiseless coding of correlated information sources*. Information Theory, IEEE Transactions on, IT-19(4):471 – 480, 1973. 33
- [tB01a] Brink, Stephan ten: *Code characteristic matching for iterative decoding of serially concatenated codes*. Annales des télécommunications, pages 394–408, 2001. 10
- [tB01b] Brink, Stephan ten: *Convergence behavior of iteratively decoded parallel concatenated codes*. IEEE Transactions on Communications, 49(10):1727–1737, 2001. 10
- [TP10] Tong, Jun and Ping, Li: *Performance Analysis of Superposition Coded Modulation*. Physical Communication, 3(3):147–155, sep 2010, ISSN 18744907. 15, 50

- [TPB99] Tishby, Naftaly, Pereira, Fernando C., and Bialek, William: *The information bottleneck method*. Proc. 37th Ann. Allerton Conf. on Communications, Control, and Computing, 24(3):3–6, 1999. 2, 69, 70, 71, 72, 112
- [Ung82] Ungerboeck, G.: *Channel Coding with Multilevel/Phase Signals*. IEEE Transactions on Information Theory, 28:55–67, 1982, ISSN 0018-9448. 14
- [vdM71] Meulen, Edward C van der: *Three-Terminal Communication Channels*. Advances in Applied Probability, 3(1):pp. 120–154, 1971, ISSN 00018678. 1
- [vdM77] Meulen, Edward C. van der: *A Survey of Multi-Way Channels in Information Theory: 1961-1976*. IEEE Transactions on Information Theory, IT-23(1):1–37, 1977. 1
- [Vit63] Viterbi, A.J.: *Phase-locked loop dynamics in the presence of noise by Fokker-Planck techniques*. Proceedings of the IEEE, 51(12):1737–1753, 1963, ISSN 0018-9219. 56, 57
- [Wei12] Weitkemper, Petra: *Signalverarbeitung für die Relay-gestützte Datenübertragung*. PhD thesis, 2012. 1, 2, 27, 30, 31, 35, 36, 37, 38, 111
- [WFH99] Wachsmann, Udo, Fischer, Robert, and Huber, Johannes B.: *Multilevel codes: Theoretical concepts and practical design rules*. IEEE Transactions on Information Theory, 45(5):1361–1391, jul 1999, ISSN 00189448. 14
- [WH10a] Wang, Li and Hanzo, Lajos: *Optimum Time Resource Allocation for TDMA-Based Differential Decode-and-Forward Cooperative Systems: A Capacity Perspective*. IEEE Communications Letters, 14(6):506–508, 2010. 1
- [WH10b] Wo, Tianbin and Hoeher, Peter Adam: *Superposition Mapping with Application in Bit-Interleaved Coded Modulation*. International ITG Conference on Source and Channel Coding, pages 1–6, 2010. 14, 15
- [Win14] Winkelbauer, Andreas: *Blind performance estimation and quantizer design with applications to relay networks*. Dissertation, 2014. 71, 73, 141
- [Wo11] Wo, Tianbin: *Superposition Mapping & Related Coding Techniques*. 2011. 15, 16, 17, 18, 21, 51, 111
- [Wyn78] Wyner, A. D.: *The Rate-Distortion Function for Source Coding with Side Information at the Decoder-II: General Sources*. 80:60–80, 1978. 33
- [WZ76] Wyner, A. and Ziv, J.: *The rate-distortion function for source coding with side information at the decoder*. IEEE Transactions on Information Theory, 22(1):1–10, 1976, ISSN 0018-9448. 33

- [ZBBW10] Zeitler, Georg, Brehmer, Johannes, Bauch, Gerhard, and Widmer, Joerg: *Source coding rate allocation in orthogonal compress-and-forward relay networks*. IEEE International Conference on Communications, (Icc), 2010, ISSN 05361486. 69
- [ZBW12] Zeitler, Georg, Bauch, Gerhard, and Widmer, Jörg: *Quantize-and-Forward Schemes for the Orthogonal Multiple-Access Relay Channel*. IEEE TRANSACTIONS ON COMMUNICATIONS, 60(4):62–75, 2012. 69, 70, 112
- [Zei12] Zeitler, Georg Christoph: *Low-Precision Quantizer Design for Communication Problems*. Dissertation, 2012. 2, 33, 69, 70, 71, 74, 75, 76, 112, 141
- [ZKBW08] Zeitler, Georg, Koetter, Ralf, Bauch, Gerhard, and Widmer, Joerg: *Design of network coding functions in multihop relay networks*. Turbo Codes and Related Topics, 2008 5th International Symposium on, pages 249–254, 2008. 69
- [ZKBW09a] Zeitler, Georg, Koetter, Ralf, Bauch, G., and Widmer, J.: *On quantizer design for soft values in the multiple-access relay channel*. In *Communications, 2009. ICC'09. IEEE International Conference on*, number Icc, pages 1–5, 2009, ISBN 9781424434350. 69, 70, 112
- [ZKBW09b] Zeitler, Georg, Koetter, Ralf, Bauch, Gerhard, and Widmer, Joerg: *An adaptive compress-and-forward scheme for the orthogonal multiple-access relay channel*. 2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, pages 1838–1842, 2009. 69

Appendix

A Entropy and Mutual Information

The statistical measures entropy and mutual information are basic concepts of information theory. The concept of entropy is a measure of uncertainty either for a single random variable or multiple variables, whereby the latter case is distinguished into joint and conditional entropy. In addition, entropy is defined on discrete random variables, while the same concept is known as differential entropy on continuous variables. Mutual information is the information that two or more random variables share and can be considered as the reduction of uncertainty of a random variable due to knowledge about the other. Proofs and detailed derivation of the following definitions can be found in [CT91].

A.1 Entropy

Suppose X is a discrete random variable with probability mass function (pmf) $\Pr\{x\}$ and alphabet $x \in \mathcal{X}$. Then, the entropy of X is defined as

$$\begin{aligned} H(X) &= -\mathbb{E} \{\log_2 \Pr\{x\}\} \\ &= - \sum_{x \in \mathcal{X}} \Pr\{x\} \log_2 \Pr\{x\} \end{aligned} \tag{A.1}$$

As the logarithm is of base 2, the unit of entropy is bits. Entropy only depends on the statistics of X and is literally the negative expectation of the logarithm of its pmf. Maximum entropy, which is the highest uncertainty about X , is achieved for a uniform distribution. In this context, the uncertainty is a measure about the average number of bits that are needed to describe the random variable. That is why a uniformly distributed random variable cannot be compressed without loss.

A.2 Joint and Conditional Entropy

The joint entropy of a pair of random variables is defined in a similar way as for a single one.

$$\begin{aligned} H(X, Y) &= -E \{ \log_2 \Pr\{x, y\} \} \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \Pr\{x, y\} \log_2 \Pr\{x, y\} \end{aligned} \quad (\text{A.2})$$

For the entropy of Y given X , again the negative expected value of logarithm of the corresponding pmf is taken.

$$\begin{aligned} H(Y|X) &= -E \{ \log_2 \Pr\{y|x\} \} \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \Pr\{x, y\} \log_2 \Pr\{y|x\} \end{aligned} \quad (\text{A.3})$$

From these two quantities, the chain rule of entropies appears to be

$$H(X, Y) = H(Y|X) + H(X) \quad (\text{A.4})$$

simply by using logarithmic law $\log_2 \Pr\{x, y\} = \log_2 \Pr\{y|x\} + \log_2 \Pr\{x\}$ and taking expectation of both sides. Generalization to more than two random variables yields

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i | X_1, X_2, \dots, X_{i-1}). \quad (\text{A.5})$$

A.3 Mutual Information for Discrete Random Variables

As illustrated in Fig. A.1, mutual information $I(X; Y)$ denotes the information that one random variable X has about another Y and vice versa, that is, it is symmetric. Mathematically, its definition is

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \Pr\{x, y\} \log_2 \frac{\Pr\{x, y\}}{\Pr\{x\} \Pr\{y\}} \quad (\text{A.6})$$

where the knowledge of Y or X reduces the uncertainty about X or Y , respectively. Obviously, the mutual information $I(X; X) = H(X) - H(X|X) = H(X)$ of a random variable with itself is the entropy of that variable. Hence, $H(X)$ is the mean information of the source, from which the uncertainty $H(X|Y)$ (also known as equivocation) of the source given the observed output has to be subtracted. Due to mutuality, one can also subtract the irrelevance $H(Y|X)$ from the mean information $H(X)$ of the source which contains the irrelevant noise. Extending (A.6) by condition

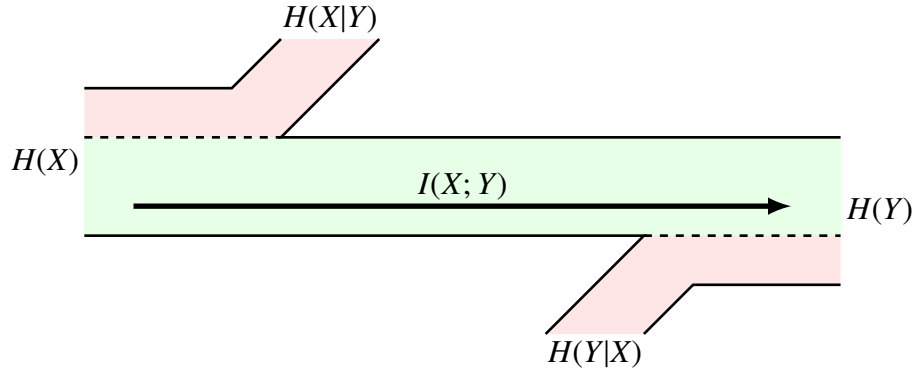


Figure A.1: Visualization of mutual information is difference of entropies.

on Z delivers the conditional mutual information.

$$I(X; Y|Z) = H(X|Z) - H(X|Y, Z) \quad (\text{A.7a})$$

$$= H(Y|Z) - H(Y|X, Z) \quad (\text{A.7b})$$

$$= H(Y|Z) + H(X|Z) - H(X, Y|Z) \quad (\text{A.7c})$$

$$= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \Pr\{x, y, z\} \log_2 \frac{\Pr\{x, y|z\}}{\Pr\{x|z\} \Pr\{y|z\}} \quad (\text{A.7d})$$

Finally, ensuing from the chain rule of entropies, the chain rule for mutual information reads as follows.

$$I(X_1, X_2, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y|X_1, X_2, \dots, X_{i-1}) \quad (\text{A.8})$$

A.4 Differential Entropy

The differential entropy for continuous random variables is very similar defined as the entropy of discrete random variables. As continuous random variables, e.g., X with probability density function (pdf) $p(x)$ generally have an unlimited support, e.g., \mathcal{X} , the sums in (A.1) change to integrals to obtain differential entropy.

$$h(X) = - \int_{\mathcal{X}} p(x) \log_2 p(x) dx \quad (\text{A.9})$$

It is important to note that in contrary to the entropy for discrete random variables, differential entropy can be negative and may also not exist, e.g., if the integral or a pdf do not exist. Similarly as before, joint and conditional differential entropies are defined as

$$h(X_1, X_2, \dots, X_n) = \int p(x_1, x_2, \dots, x_n) \log_2 p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \quad (\text{A.10})$$

and

$$h(X_1|X_2, \dots, X_n) = \int p(x_1, x_2, \dots, x_n) \log_2 p(x_1|x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \quad (\text{A.11})$$

respectively. The corresponding chain rule is

$$h(X_1, X_2, \dots, X_n) = \sum_{i=1}^n h(X_i|X_1, X_2, \dots, X_{i-1}). \quad (\text{A.12})$$

A.5 Relative Entropy

The concept of relative entropy or Kullback-Leibler distance is a measure of the similarity of two distribution functions. In other words, it is the inefficiency of describing a random variable X with pdf $q(x)$ instead of the true pdf $p(x)$.

$$D_{KL}(p||q) = \int_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)} dx \quad (\text{A.13})$$

The relative entropy $D_{KL}(p||q)$ denotes then the number of additional Bits that are necessary to describe x when using $q(x)$ instead of $p(x)$. The same principle can also be applied to discrete random variables by substituting integrals by sums. Moreover, the concept can be extended easily to joint or conditioned pdfs or pmfs. In Ch. 6 for example the following conditioned Kullback-Leibler distance is required.

$$D_{KL}(\Pr\{x|y\}||\Pr\{x|z\}) = \sum_{x \in \mathcal{X}} \Pr\{x|y\} \log_2 \frac{\Pr\{x|y\}}{\Pr\{x|z\}} \quad (\text{A.14})$$

A.6 Mutual Information for Continuous Variables

The (conditional) mutual information for continuous variables and its chain rule are defined in the same way as for discrete variables: Simply by replacing entropies by differential entropies.

$$I(X; Y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dx dy \quad (\text{A.15a})$$

A.7 Differential Entropy of Normal Distribution

The differential entropy of a normal distribution is quite important in information theory because it maximizes differential entropy given an average power constraint, i.e., variance limited. For a random variable $X \sim p(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{x^2}{2\sigma_X^2}}$, the differential entropy using natural logarithm is

$$h(X) = \frac{1}{2} \log_2(2\pi e \sigma_X^2) \quad (\text{A.16})$$

Following a similar derivation, the differential entropy of multivariate normal distributed variables $X_1, \dots, X_n \sim \mathcal{N}_n(\mu, \mathbf{R})$ becomes

$$h(X_1, \dots, X_n) = \frac{1}{2} \log_2((2\pi e)^n \det \mathbf{R}), \quad (\text{A.17})$$

whereby a complex Gaussian denotes a special case: Let N be a complex Gaussian random variable denoting AWGN, i.e., a realization n follows $\mathcal{CN}(0, \sigma_N^2)$ which is equivalent to $\mathcal{N}_2(0, \mathbf{R}_{NN})$ with $\mathbf{R}_{NN} = \frac{1}{2}\sigma_N^2 \mathbf{I}$ (\mathbf{I} denotes identity matrix).

$$h(N) = \log_2 \pi e \sigma_N^2 \quad (\text{A.18})$$

B Random Binning as in Wyner-Ziv Coding

A general explanation of random binning for Wyner-Ziv coding can be found in [EK12]. Here random binning will be explained for CF as introduced in Sec. 3.3.3. The relay has received y_R according to (3.15a) and compresses it to \hat{y}_R while the destination has already received y_{D1} according to (3.15b) which is correlated to \hat{y}_R . As the focus is on the concept of random binning, let us assume a strongly simplified toy example where x_{S1} is BPSK modulated and the compression $y_R \rightarrow \hat{y}_R$ is a simple uniform quantization with 4 levels $\hat{\mathcal{Y}}_R = \{1, 2, 3, 4\}$ (cf. Fig. A.2) which is typically not optimal (cf. Ch. 6).¹ Usually, the quantized signal $\hat{y}_R \in \{1, 2, 3, 4\}$ would have the rate $\tilde{r} = H(\hat{\mathcal{Y}}_R)$ but, due to the side information Y_{D1} at the destination, this rate can be lowered to $r = H(\hat{\mathcal{Y}}_R | Y_{D1})$ with the help of random binning (Wyner-Ziv coding). Please note that knowledge about Y_{D1} is only required at the Wyner-Ziv source decoder. Table B.1 shows an arbitrary binning to visualize the rate reduction exploiting side information. In particular, the quantization indices $\hat{y}_R = 1$ and $\hat{y}_R = 4$ are grouped into the same bin B with index $b = 1$ while $\hat{y}_R = 2$ and $\hat{y}_R = 3$ fall into $B(b = 2)$. Forwarding $b \in \{1, 2\}$ is obviously possible with 1 Bit

¹Please note that \hat{y}_R is not a representative but only a compression index which has to be appropriately interpreted, e.g., with the probability $\Pr\{x_{S1} | \hat{y}_R\}$ as in Ch. 6.

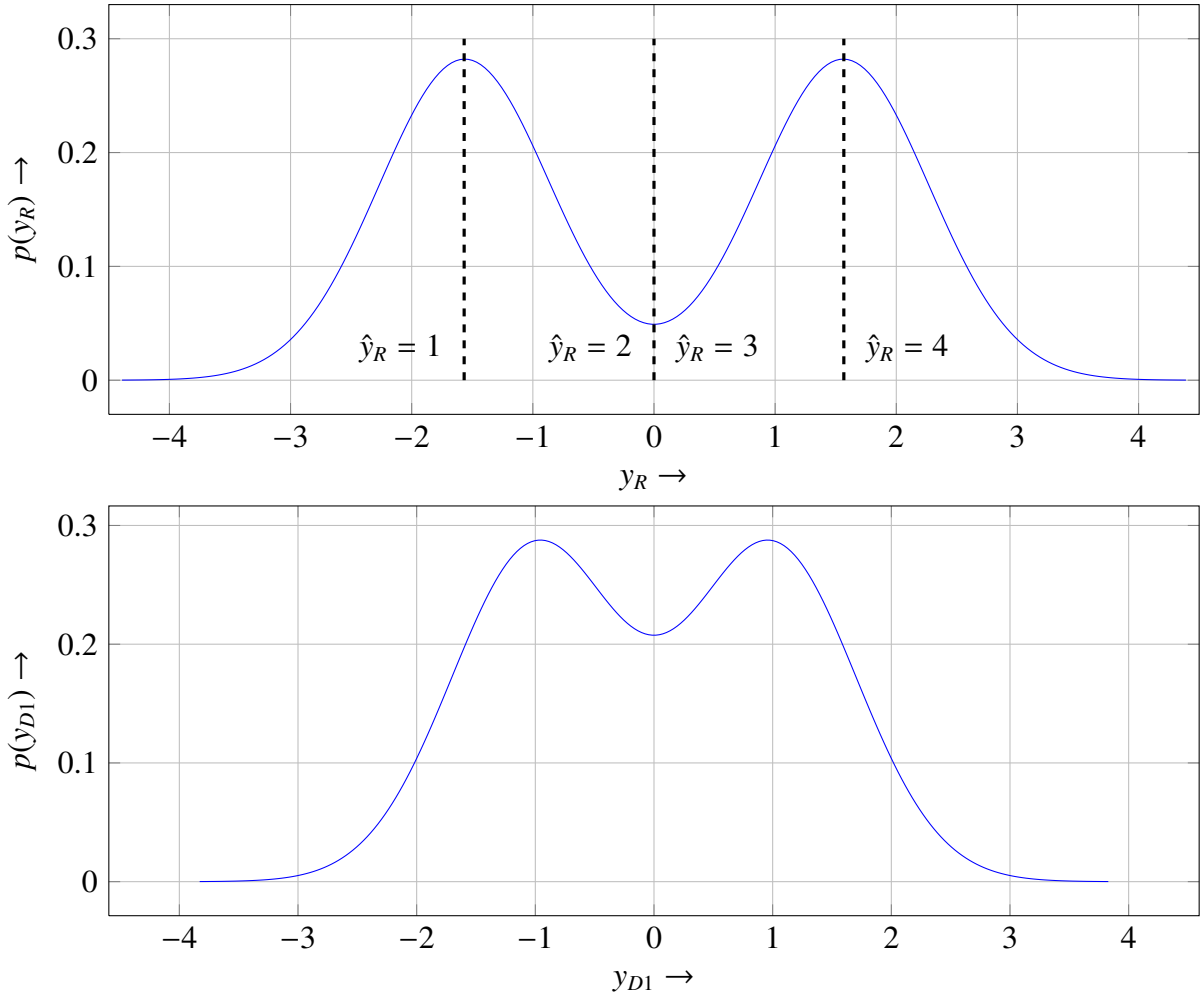


Figure A.2: pdfs of y_R and y_{D1} for BPSK input with SNRs (3.9, 0dB) according to a line setup: Source, relay, and destination are at positions $d_S = 0$, $d_R = 0.8$, and $d_D = 1$ with pathloss exponent $\alpha = 4$, transmit power $P_{S1} = 1$ and noise power $N_0 = 1$.

which is less than $H(\hat{Y}_R) \approx 2$ (\hat{y}_R is almost uniformly distributed). The decoding at the destination is very simple: Depending on $y_{D1} > 0$ or $y_{D1} < 0$ and $b = 1$ has been received, \hat{y}_R is either 1 or 4, respectively. That is why all \hat{y}_R have been grouped in a mirrored fashion. Please note that the scheme in Table B.1 is not optimal (and may cause an information loss) but serves only the purpose of illustration. In fact, a whole vector of length o of indices \hat{y}_R is entropy-encoded with rate \tilde{r} to $2^{o\tilde{r}}$ sequences with index s . Afterwards, these indices $s \in \{1, 2, \dots, 2^{oR}\}$ are uniformly distributed among 2^{Mr} bins $\mathcal{B}(b)$ where $r \leq \tilde{r}$.

Table B.1: Simple Wyner-Ziv coding scheme using random binning.

	$\mathcal{B}(b = 1)$		$\mathcal{B}(b = 2)$	
	$\hat{y}_R = 1$	$\hat{y}_R = 4$	$\hat{y}_R = 2$	$\hat{y}_R = 3$
$y_{D1} > 0$	•		•	
$y_{D1} < 0$		•		•

C Achievable Rates for Discrete Input

C.1 AWGN Channel

As digital communication systems use discrete modulation alphabets, the achievable rate of an AWGN channel for a specific discrete input is considered. Therefore, mutual information is split into the difference of input entropy and conditional input entropy.

$$I(X; Y) = H(X) - H(X|Y) \quad (\text{C.19})$$

For common digital modulations schemes, $H(X)$ can be very easily determined especially for uniform distributions. The second term however has to be evaluated numerically, e.g., by randomly generating realizations of x and y according to (2.1) and calculating

$$\Pr\{x|y\} = \frac{p(y|x) \Pr\{x\}}{\sum_{x \in \mathcal{X}} p(y|x) \Pr\{x\}}, \quad (\text{C.20})$$

where $p(y|x)$ is a shifted Gaussian $CN(h\sqrt{P}x, 1)$ and the denominator nothing else than normalization.

C.2 Max Flow Min Cut Sets for Relay Channel

The achievable rate for DF (cf. (3.13)) can only be solved analytically for the Gaussian relay channel which implicates a Gaussian distributed input (cf. (3.28)). For a practical scheme with

digital modulation however, the input alphabet $\Pr\{x_{S1}, x_{S2}, x_R\}$ is always discrete and, hence, (3.13) can only be solved numerically. According to (3.13), 4 information terms are of interest, where $I(X_{S1}; Y_R)$ and $I(X_{S1}; Y_{D1})$ are Point-to-Point (P2P) transmissions as in Sec. 2.2 and $I(X_{S2}; Y_{D2}^{non}|X_R)$ is, according to chain rule, included in

$$\begin{aligned} I(X_{S2}, X_R; Y_{D2}^{non}) &= I(X_R; Y_{D2}^{non}) + I(X_{S2}; Y_{D2}^{non}|X_R) \\ &= E \left\{ \log_2 \frac{\Pr\{x_R|y_{D2}^{non}\}}{\Pr\{x_R\}} \right\} + E \left\{ \log_2 \frac{\Pr\{x_{S2}|x_R, y_{D2}^{non}\}}{\Pr\{x_{S2}|x_R\}} \right\}. \end{aligned} \quad (\text{C.21a})$$

A priori probabilities as well as entropies in the denominators are straightforward with $\Pr\{x_{S2}|x_R\} = \Pr\{\tilde{x}_{S2}\}$. The a posteriori probabilities

$$\Pr\{x_R|y_{D2}^{non}\} = \frac{\sum_{\tilde{x}_{S2} \in \tilde{\mathcal{X}}_{S2}} p(y_{D2}^{non}|x_R, \tilde{x}_{S2}) \Pr\{x_R, \tilde{x}_{S2}\}}{\sum_{x_R \in \mathcal{X}_R} \sum_{\tilde{x}_{S2} \in \tilde{\mathcal{X}}_{S2}} p(y_{D2}^{non}|x_R, \tilde{x}_{S2}) \Pr\{x_R, \tilde{x}_{S2}\}} \quad (\text{C.21b})$$

and

$$\Pr\{x_{S2}|x_R, y_{D2}^{non}\} = \frac{p(y_{D2}^{non}|x_R, \tilde{x}_{S2}) \Pr\{x_R, \tilde{x}_{S2}\}}{\sum_{\tilde{x}_{S2} \in \tilde{\mathcal{X}}_{S2}} p(y_{D2}^{non}|x_R, \tilde{x}_{S2}) \Pr\{x_R, \tilde{x}_{S2}\}} \quad (\text{C.21c})$$

are determined by numerical evaluation whereby $p(y_{D2}^{non}|x_R, \tilde{x}_{S2})$ is usually a shifted noise distribution, $CN(h_{SD} \sqrt{P_{S2}} \sqrt{1 - \rho^2} \cdot \tilde{x}_{S2} + (h_{SD} \sqrt{P_{S2}} \cdot \rho + h_{RD} \sqrt{P_R}) \cdot x_R, 1)$. Hereby, an unusual case is the presence of phase errors when applying distributed beamforming for DF which is discussed in Ch. 5.

D Log-Likelihood-Ratios in a Turbo Decoder

Following the symbol by symbol Maximum-A-Posteriori (MAP) criterion the Log Likelihood Ratio (LLR)

$$L(u_l) = \log \frac{\Pr\{u_l = 0, \mathbf{y}\}}{\Pr\{u_l = 1, \mathbf{y}\}} \quad (\text{D.22a})$$

is of interest, whereby the occurring joint probabilities are not directly accessible. Therefore, the set Γ of all possible code words \mathbf{c} is divided into subsets $\Gamma_l^{(0)}$ and $\Gamma_l^{(1)}$ of code words whose l th information bit is $u_l = 0$ and $u_l = 1$, respectively.

$$L(u_l) = \log \frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \Pr\{\mathbf{y}|\mathbf{c}\} \Pr\{\mathbf{c}\}}{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \Pr\{\mathbf{y}|\mathbf{c}\} \Pr\{\mathbf{c}\}} \quad (\text{D.22b})$$

Exploiting the fact that elements y_i of \mathbf{y} are independent when \mathbf{c} is known and $\Pr\{\mathbf{c}\} = \Pr\{\mathbf{u}\}$ due to bijectivity, where elements u_j of \mathbf{u} are independent as well, leads to

$$L(u_l) = \log \frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \prod_{i=0}^{n-1} \Pr\{y_i|c_i\} \prod_{j=0}^{k-1} \Pr\{u_j\}}{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \prod_{i=0}^{n-1} \Pr\{y_i|c_i\} \prod_{j=0}^{k-1} \Pr\{u_j\}} \quad (\text{D.22c})$$

Finally, as the component encoders are systematic ($u_l = c_l$), $\Pr\{y_l|c_l\} = \Pr\{y_l|u_l\}$ and $\Pr\{u_l\}$ can be extracted as constant factors from the sum and split up due to logarithmic laws.

$$L(u_l) = \log \frac{\Pr\{y_l|u_l = 0\}}{\Pr\{y_l|u_l = 1\}} + \log \frac{\Pr\{u_l = 0\}}{\Pr\{u_l = 1\}} + \log \frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \prod_{i=0, i \neq l}^{n-1} \Pr\{y_i|c_i\} \prod_{j=0, j \neq l}^{k-1} \Pr\{u_j\}}{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \prod_{i=0, i \neq l}^{n-1} \Pr\{y_i|c_i\} \prod_{j=0, j \neq l}^{k-1} \Pr\{u_j\}} \quad (\text{D.22d})$$

$$= L(y_l|u_l) + L_a(u_l) + \log \underbrace{\frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \prod_{i=0, i \neq l}^{n-1} e^{-(L(y_i|c_i))c_i} \prod_{j=0, j \neq l}^{k-1} e^{-L(u_j)u_j}}{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \prod_{i=0, i \neq l}^{n-1} e^{-(L(y_i|c_i))c_i} \prod_{j=0, j \neq l}^{k-1} e^{-L(u_j)u_j}}}_{L_e(u_l)} \quad (\text{D.22e})$$

E Superposition Multilevel Coding - ρ for Parallel Detection

The following tables depict the possible values for ρ for different parameters K , L , and G of a superposition multilevel code with joint demapping. The scenario is assumed to be a line setup with source, relay, and destination at positions 0, d , and 1, i.e., only the relay position is variable and determines all channel (attenuation) coefficients of the relay channel. All average transmit powers are set to unity.

E.1 Complex Group Size $G = 6$:Table E.2: Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.3$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 9$	0.521	0.298	0.242	0.229	0.225	0.224
$K = 10$	0.685	0.464	0.410	0.397	0.394	0.393
$K = 11$	0.844	0.623	0.570	0.556	0.553	0.552
$K = 12$	1.000	0.777	0.723	0.710	0.706	0.705
$K = 13$	-	0.814	0.761	0.747	0.744	0.743
$K = 14$	-	0.852	0.798	0.785	0.781	0.780
$K = 15$	-	0.889	0.835	0.822	0.818	0.818
$K = 16$	-	0.926	0.872	0.859	0.855	0.854
$K = 17$	-	0.963	0.909	0.895	0.892	0.891
$K = 18$	-	1.000	0.945	0.932	0.929	0.928
$K = 19$	-	-	0.955	0.941	0.938	0.937
$K = 20$	-	-	0.964	0.950	0.947	0.946
$K = 21$	-	-	0.973	0.959	0.956	0.955
$K = 22$	-	-	0.982	0.968	0.965	0.964
$K = 23$	-	-	0.991	0.977	0.974	0.973
$K = 24$	-	-	1.000	0.986	0.983	0.982
$K = 25$	-	-	-	0.989	0.985	0.984
$K = 26$	-	-	-	0.991	0.988	0.987
$K = 27$	-	-	-	0.993	0.990	0.989
$K = 28$	-	-	-	0.995	0.992	0.991
$K = 29$	-	-	-	0.998	0.994	0.994
$K = 30$	-	-	-	1.000	0.997	0.996
$K = 31$	-	-	-	-	0.997	0.996
$K = 32$	-	-	-	-	0.998	0.997
$K = 33$	-	-	-	-	0.998	0.997
$K = 34$	-	-	-	-	0.999	0.998
$K = 35$	-	-	-	-	0.999	0.999
$K = 36$	-	-	-	-	1.000	0.999

Table E.3: Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.4$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 9$	0.176	-	-	-	-	-
$K = 10$	0.488	0.047	-	-	-	-
$K = 11$	0.755	0.377	0.275	0.249	0.242	0.241
$K = 12$	1.000	0.644	0.554	0.531	0.525	0.524
$K = 13$	-	0.706	0.617	0.595	0.589	0.588
$K = 14$	-	0.767	0.679	0.657	0.652	0.650
$K = 15$	-	0.827	0.740	0.718	0.713	0.711
$K = 16$	-	0.885	0.799	0.778	0.772	0.771
$K = 17$	-	0.943	0.858	0.836	0.831	0.830
$K = 18$	-	1.000	0.915	0.894	0.889	0.887
$K = 19$	-	-	0.929	0.908	0.903	0.902
$K = 20$	-	-	0.944	0.923	0.917	0.916
$K = 21$	-	-	0.958	0.937	0.932	0.930
$K = 22$	-	-	0.972	0.951	0.946	0.944
$K = 23$	-	-	0.986	0.965	0.960	0.958
$K = 24$	-	-	1.000	0.979	0.974	0.972
$K = 25$	-	-	-	0.983	0.977	0.976
$K = 26$	-	-	-	0.986	0.981	0.979
$K = 27$	-	-	-	0.990	0.984	0.983
$K = 28$	-	-	-	0.993	0.988	0.986
$K = 29$	-	-	-	0.997	0.991	0.990
$K = 30$	-	-	-	1.000	0.995	0.993

Table E.4: Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.5$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 9$	-	-	-	-	-	-
$K = 10$	-	-	-	-	-	-
$K = 11$	0.538	-	-	-	-	-
$K = 12$	1.000	0.286	0.000	-	-	-
$K = 13$	-	0.433	0.214	0.149	0.131	0.127
$K = 14$	-	0.563	0.372	0.319	0.305	0.302
$K = 15$	-	0.682	0.507	0.460	0.448	0.445
$K = 16$	-	0.793	0.629	0.586	0.575	0.572
$K = 17$	-	0.899	0.742	0.701	0.691	0.688
$K = 18$	-	1.000	0.849	0.810	0.800	0.798
$K = 19$	-	-	0.874	0.836	0.826	0.824
$K = 20$	-	-	0.900	0.862	0.852	0.850
$K = 21$	-	-	0.925	0.888	0.878	0.876
$K = 22$	-	-	0.951	0.913	0.904	0.901
$K = 23$	-	-	0.975	0.938	0.929	0.927
$K = 24$	-	-	1.000	0.963	0.954	0.952
$K = 25$	-	-	-	0.969	0.960	0.958
$K = 26$	-	-	-	0.975	0.966	0.964
$K = 27$	-	-	-	0.982	0.972	0.970
$K = 28$	-	-	-	0.988	0.979	0.976
$K = 29$	-	-	-	0.994	0.985	0.982
$K = 30$	-	-	-	1.000	0.991	0.989
$K = 31$	-	-	-	-	0.992	0.990
$K = 32$	-	-	-	-	0.994	0.992
$K = 33$	-	-	-	-	0.995	0.993
$K = 34$	-	-	-	-	0.997	0.995
$K = 35$	-	-	-	-	0.998	0.996
$K = 36$	-	-	-	-	1.000	0.998

Table E.5: Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.6$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 9$	-	-	-	-	-	-
$K = 10$	-	-	-	-	-	-
$K = 11$	-	-	-	-	-	-
$K = 12$	1.000	-	-	-	-	-
$K = 13$	-	-	-	-	-	-
$K = 14$	-	-	-	-	-	-
$K = 15$	-	0.112	-	-	-	-
$K = 16$	-	0.512	-	-	-	-
$K = 17$	-	0.778	0.356	0.204	0.158	0.145
$K = 18$	-	1.000	0.657	0.557	0.531	0.524
$K = 19$	-	-	0.720	0.626	0.601	0.595
$K = 20$	-	-	0.781	0.690	0.667	0.661
$K = 21$	-	-	0.838	0.752	0.729	0.723
$K = 22$	-	-	0.894	0.810	0.789	0.783
$K = 23$	-	-	0.948	0.867	0.846	0.841
$K = 24$	-	-	1.000	0.921	0.901	0.896
$K = 25$	-	-	-	0.935	0.915	0.910
$K = 26$	-	-	-	0.948	0.928	0.923
$K = 27$	-	-	-	0.961	0.941	0.936
$K = 28$	-	-	-	0.974	0.955	0.950
$K = 29$	-	-	-	0.987	0.968	0.963
$K = 30$	-	-	-	1.000	0.981	0.976
$K = 31$	-	-	-	-	0.984	0.979
$K = 32$	-	-	-	-	0.987	0.982
$K = 33$	-	-	-	-	0.990	0.986
$K = 34$	-	-	-	-	0.994	0.989
$K = 35$	-	-	-	-	0.997	0.992
$K = 36$	-	-	-	-	1.000	0.995

Table E.6: Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.7$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 9$	-	-	-	-	-	-
$K = 10$	-	-	-	-	-	-
$K = 11$	-	-	-	-	-	-
$K = 12$	1.000	-	-	-	-	-
$K = 13$	-	-	-	-	-	-
$K = 14$	-	-	-	-	-	-
$K = 15$	-	-	-	-	-	-
$K = 16$	-	-	-	-	-	-
$K = 17$	-	0.053	-	-	-	-
$K = 18$	-	1.000	-	-	-	-
$K = 19$	-	-	-	-	-	-
$K = 20$	-	-	0.090	-	-	-
$K = 21$	-	-	0.456	-	-	-
$K = 22$	-	-	0.673	0.317	0.171	0.120
$K = 23$	-	-	0.848	0.574	0.490	0.467
$K = 24$	-	-	1.000	0.765	0.698	0.681
$K = 25$	-	-	-	0.807	0.743	0.727
$K = 26$	-	-	-	0.848	0.787	0.771
$K = 27$	-	-	-	0.888	0.828	0.813
$K = 28$	-	-	-	0.926	0.868	0.854
$K = 29$	-	-	-	0.964	0.907	0.893
$K = 30$	-	-	-	1.000	0.945	0.931
$K = 31$	-	-	-	-	0.954	0.941
$K = 32$	-	-	-	-	0.964	0.950
$K = 33$	-	-	-	-	0.973	0.959
$K = 34$	-	-	-	-	0.982	0.968
$K = 35$	-	-	-	-	0.991	0.977
$K = 36$	-	-	-	-	1.000	0.987

Table E.7: Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.8$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 9$	-	-	-	-	-	-
$K = 10$	-	-	-	-	-	-
$K = 11$	-	-	-	-	-	-
$K = 12$	1.000	-	-	-	-	-
$K = 13$	-	-	-	-	-	-
$K = 14$	-	-	-	-	-	-
$K = 15$	-	-	-	-	-	-
$K = 16$	-	-	-	-	-	-
$K = 17$	-	-	-	-	-	-
$K = 18$	-	1.000	-	-	-	-
$K = 19$	-	-	-	-	-	-
$K = 20$	-	-	-	-	-	-
$K = 21$	-	-	-	-	-	-
$K = 22$	-	-	-	-	-	-
$K = 23$	-	-	-	-	-	-
$K = 24$	-	-	1.000	-	-	-
$K = 25$	-	-	-	-	-	-
$K = 26$	-	-	-	-	-	-
$K = 27$	-	-	-	0.252	-	-
$K = 28$	-	-	-	0.599	-	-
$K = 29$	-	-	-	0.821	0.456	0.315
$K = 30$	-	-	-	1.000	0.718	0.631
$K = 31$	-	-	-	-	0.771	0.690
$K = 32$	-	-	-	-	0.821	0.745
$K = 33$	-	-	-	-	0.869	0.796
$K = 34$	-	-	-	-	0.914	0.845
$K = 35$	-	-	-	-	0.958	0.892
$K = 36$	-	-	-	-	1.000	0.936

Table E.8: Values of ρ for different combinations of K and L for $G = 6$ and $d = 0.9$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 9$	-	-	-	-	-	-
$K = 10$	-	-	-	-	-	-
$K = 11$	-	-	-	-	-	-
$K = 12$	1.000	-	-	-	-	-
$K = 13$	-	-	-	-	-	-
$K = 14$	-	-	-	-	-	-
$K = 15$	-	-	-	-	-	-
$K = 16$	-	-	-	-	-	-
$K = 17$	-	-	-	-	-	-
$K = 18$	-	1.000	-	-	-	-
$K = 19$	-	-	-	-	-	-
$K = 20$	-	-	-	-	-	-
$K = 21$	-	-	-	-	-	-
$K = 22$	-	-	-	-	-	-
$K = 23$	-	-	-	-	-	-
$K = 24$	-	-	1.000	-	-	-
$K = 25$	-	-	-	-	-	-
$K = 26$	-	-	-	-	-	-
$K = 27$	-	-	-	-	-	-
$K = 28$	-	-	-	-	-	-
$K = 29$	-	-	-	-	-	-
$K = 30$	-	-	-	1.000	-	-
$K = 31$	-	-	-	-	-	-
$K = 32$	-	-	-	-	-	-
$K = 33$	-	-	-	-	-	-
$K = 34$	-	-	-	-	-	-
$K = 35$	-	-	-	-	-	-
$K = 36$	-	-	-	-	1.000	-

E.2 Complex Group Size $G = 4$:Table E.9: Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.3$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 5$	0.259	0.020	-	-	-	-
$K = 6$	0.521	0.298	0.242	0.229	0.225	0.224
$K = 7$	0.765	0.544	0.491	0.478	0.474	0.473
$K = 8$	1.000	0.777	0.723	0.710	0.706	0.705
$K = 9$	-	0.833	0.779	0.766	0.763	0.762
$K = 10$	-	0.889	0.835	0.822	0.818	0.818
$K = 11$	-	0.945	0.890	0.877	0.874	0.873
$K = 12$	-	1.000	0.945	0.932	0.929	0.928
$K = 13$	-	-	0.959	0.946	0.942	0.941
$K = 14$	-	-	0.973	0.959	0.956	0.955
$K = 15$	-	-	0.986	0.973	0.969	0.969
$K = 16$	-	-	1.000	0.986	0.983	0.982
$K = 17$	-	-	-	0.990	0.986	0.986
$K = 18$	-	-	-	0.993	0.990	0.989
$K = 19$	-	-	-	0.997	0.993	0.992
$K = 20$	-	-	-	1.000	0.997	0.996

Table E.10: Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.4$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 5$	-	-	-	-	-	-
$K = 6$	0.176	-	-	-	-	-
$K = 7$	0.625	0.225	0.110	0.080	0.072	0.070
$K = 8$	1.000	0.644	0.554	0.531	0.525	0.524
$K = 9$	-	0.737	0.648	0.626	0.621	0.619
$K = 10$	-	0.827	0.740	0.718	0.713	0.711
$K = 11$	-	0.914	0.829	0.807	0.802	0.801
$K = 12$	-	1.000	0.915	0.894	0.889	0.887
$K = 13$	-	-	0.937	0.915	0.910	0.909
$K = 14$	-	-	0.958	0.937	0.932	0.930
$K = 15$	-	-	0.979	0.958	0.953	0.951
$K = 16$	-	-	1.000	0.979	0.974	0.972
$K = 17$	-	-	-	0.984	0.979	0.978
$K = 18$	-	-	-	0.990	0.984	0.983
$K = 19$	-	-	-	0.995	0.990	0.988
$K = 20$	-	-	-	1.000	0.995	0.993

Table E.11: Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.5$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 5$	-	-	-	-	-	-
$K = 6$	-	-	-	-	-	-
$K = 7$	0.236	-	-	-	-	-
$K = 8$	1.000	0.286	0.000	-	-	-
$K = 9$	-	0.500	0.297	0.239	0.224	0.220
$K = 10$	-	0.682	0.507	0.460	0.448	0.445
$K = 11$	-	0.847	0.686	0.644	0.634	0.631
$K = 12$	-	1.000	0.849	0.810	0.800	0.798
$K = 13$	-	-	0.887	0.849	0.839	0.837
$K = 14$	-	-	0.925	0.888	0.878	0.876
$K = 15$	-	-	0.963	0.926	0.916	0.914
$K = 16$	-	-	1.000	0.963	0.954	0.952
$K = 17$	-	-	-	0.972	0.963	0.961
$K = 18$	-	-	-	0.982	0.972	0.970
$K = 19$	-	-	-	0.991	0.982	0.979
$K = 20$	-	-	-	1.000	0.991	0.989

Table E.12: Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.6$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 5$	-	-	-	-	-	-
$K = 6$	-	-	-	-	-	-
$K = 7$	-	-	-	-	-	-
$K = 8$	1.000	-	-	-	-	-
$K = 9$	-	-	-	-	-	-
$K = 10$	-	0.112	-	-	-	-
$K = 11$	-	0.653	0.134	-	-	-
$K = 12$	-	1.000	0.657	0.557	0.531	0.524
$K = 13$	-	-	0.751	0.658	0.634	0.628
$K = 14$	-	-	0.838	0.752	0.729	0.723
$K = 15$	-	-	0.921	0.839	0.818	0.812
$K = 16$	-	-	1.000	0.921	0.901	0.896
$K = 17$	-	-	-	0.941	0.921	0.916
$K = 18$	-	-	-	0.961	0.941	0.936
$K = 19$	-	-	-	0.981	0.961	0.956
$K = 20$	-	-	-	1.000	0.981	0.976

Table E.13: Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.7$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 5$	-	-	-	-	-	-
$K = 6$	-	-	-	-	-	-
$K = 7$	-	-	-	-	-	-
$K = 8$	1.000	-	-	-	-	-
$K = 9$	-	-	-	-	-	-
$K = 10$	-	-	-	-	-	-
$K = 11$	-	-	-	-	-	-
$K = 12$	-	1.000	-	-	-	-
$K = 13$	-	-	-	-	-	-
$K = 14$	-	-	0.456	-	-	-
$K = 15$	-	-	0.764	0.459	0.357	0.328
$K = 16$	-	-	1.000	0.765	0.698	0.681
$K = 17$	-	-	-	0.828	0.765	0.749
$K = 18$	-	-	-	0.888	0.828	0.813
$K = 19$	-	-	-	0.945	0.888	0.873
$K = 20$	-	-	-	1.000	0.945	0.931

Table E.14: Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.8$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 5$	-	-	-	-	-	-
$K = 6$	-	-	-	-	-	-
$K = 7$	-	-	-	-	-	-
$K = 8$	1.000	-	-	-	-	-
$K = 9$	-	-	-	-	-	-
$K = 10$	-	-	-	-	-	-
$K = 11$	-	-	-	-	-	-
$K = 12$	-	1.000	-	-	-	-
$K = 13$	-	-	-	-	-	-
$K = 14$	-	-	-	-	-	-
$K = 15$	-	-	-	-	-	-
$K = 16$	-	-	1.000	-	-	-
$K = 17$	-	-	-	-	-	-
$K = 18$	-	-	-	0.252	-	-
$K = 19$	-	-	-	0.717	0.253	-
$K = 20$	-	-	-	1.000	0.718	0.631

Table E.15: Values of ρ for different combinations of K and L for $G = 4$ and $d = 0.9$.

	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
$K = 5$	-	-	-	-	-	-
$K = 6$	-	-	-	-	-	-
$K = 7$	-	-	-	-	-	-
$K = 8$	1.000	-	-	-	-	-
$K = 9$	-	-	-	-	-	-
$K = 10$	-	-	-	-	-	-
$K = 11$	-	-	-	-	-	-
$K = 12$	-	1.000	-	-	-	-
$K = 13$	-	-	-	-	-	-
$K = 14$	-	-	-	-	-	-
$K = 15$	-	-	-	-	-	-
$K = 16$	-	-	1.000	-	-	-
$K = 17$	-	-	-	-	-	-
$K = 18$	-	-	-	-	-	-
$K = 19$	-	-	-	-	-	-
$K = 20$	-	-	-	1.000	-	-

F Information Bottleneck Algorithms

Algorithms described in this appendix are adapted versions of the iterative IB algorithms in Sec. 6.1: The Lagrangian multiplier β , which usually trades off compression rate against preserved relevant information, is eliminated by setting $\beta \rightarrow \infty$. Hence, rate and information are maximized (lower compression) with respect to a specific cardinality of the compressed variable. In fact, this cardinality now steers the trade-off.

Input: $\Pr\{x, y\}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}, \epsilon > 0$

Output: $\Pr\{z|y\}, \Pr\{x|z\}, (I(r), r)$

```

1 initialize  $\Pr\{z|y\}^{(0)}$  according to Maximum Output Entropy (MOE) [Win14]
2  $k \leftarrow 1$ 
3  $\Pr\{z\}^{(0)} \leftarrow \sum_y \Pr\{y\} \Pr\{z|y\}^{(0)}$ 
4  $\Pr\{x|z\}^{(0)} \leftarrow \frac{1}{\Pr\{z\}^{(0)}} \sum_y \Pr\{x, y\} \Pr\{z|y\}^{(0)}$ 
5  $d^{(0)}(z, y) \leftarrow D_{KL}(\Pr\{x|y\} \| \Pr\{x|z\}^{(0)})$ 
6 find,  $\forall y, z_y^* = \operatorname{argmin}_z d^{(0)}(z, y)$ 
7  $\Pr\{z|y\}^{(1)} \leftarrow \mathbb{1}_{z=z_y^*}$ 
8 while  $|\Pr\{z|y\}^{(k)} - \Pr\{z|y\}^{(k-1)}| / (|\mathcal{Y}| \cdot |\mathcal{Z}|) \geq \epsilon$  do
9    $\Pr\{z\}^{(k)} \leftarrow \sum_y \Pr\{y\} \Pr\{z|y\}^{(k)}$ 
10   $\Pr\{x|z\}^{(k)} \leftarrow \frac{1}{\Pr\{z\}^{(k)}} \sum_y \Pr\{x, y\} \Pr\{z|y\}^{(k)}$ 
11   $d^{(k)}(z, y) \leftarrow D_{KL}(\Pr\{x|y, y_{D1}\} \| \Pr\{x|z\}^{(k)})$ 
12  find,  $\forall y, z_y^* = \operatorname{argmin}_z d^{(k)}(z, y)$ 
13   $\Pr\{z|y\}^{(k+1)} \leftarrow \mathbb{1}_{z=z_y^*}$ 
14   $k \leftarrow k + 1$ 
15 end
16  $\Pr\{z|y\} \leftarrow \Pr\{z|y\}^{(k)}$ 
17  $\Pr\{z\} \leftarrow \sum_y \Pr\{y\} \Pr\{z|y\}$ 
18  $\Pr\{x|z\} \leftarrow \frac{1}{\Pr\{z\}} \sum_y \Pr\{x, y\} \Pr\{z|y\}$ 
19  $I(r(\beta)) \leftarrow \sum_{x,z} \Pr\{x|z\} \Pr\{z\} \log_2 \left( \frac{\Pr\{x|z\}}{\Pr\{x\}} \right)$ 
20  $r(\beta) \leftarrow \sum_{y,z} \Pr\{z|y\} \Pr\{y\} \log_2 \left( \frac{\Pr\{z|y\}}{\Pr\{z\}} \right)$ 

```

Algorithm 3: Iterative IB algorithm that delivers deterministic quantizers ($\beta \rightarrow \infty$) [Zei12], where $\mathbb{1}$ denotes the indicator function.

Input: $\Pr\{x_{S1}, y_R, y_{D1}\}, \mathcal{X}_{S1}, \mathcal{Y}_R, \mathcal{Y}_{D1}, \hat{\mathcal{Y}}_R, \epsilon > 0$

Output: $\Pr\{\hat{y}_R|y_R\}, \Pr\{x_{S1}|\hat{y}_R\}, (I(r), r)$

```

1 initialize  $\Pr\{\hat{y}_R|y_R\}^{(0)}$  according to MOE
2  $k \leftarrow 1$ 
3  $\Pr\{\hat{y}_R\}^{(0)} \leftarrow \sum_{y_R} \Pr\{y_R\} \Pr\{\hat{y}_R|y_R\}^{(0)}$ 
4  $\Pr\{\hat{y}_R, y_{D1}\}^{(0)} \leftarrow \sum_{y_R} \Pr\{y_R, y_{D1}\} \Pr\{\hat{y}_R|y_R\}^{(0)}$ 
5  $\Pr\{\hat{y}_R|y_{D1}\}^{(0)} \leftarrow \frac{\Pr\{\hat{y}_R, y_{D1}\}^{(0)}}{\Pr\{y_{D1}\}}$ 
6  $\Pr\{x_{S1}|\hat{y}_R, y_{D1}\}^{(0)} \leftarrow \frac{1}{\Pr\{\hat{y}_R, y_{D1}\}^{(0)}} \sum_{y_R} \Pr\{x_{S1}, y_R, y_{D1}\} \Pr\{\hat{y}_R|y_R\}^{(0)}$ 
7  $d^{(0)}(\hat{y}_R, y_R) \leftarrow D_{KL}(\Pr\{x_{S1}|y_R, y_{D1}\} \parallel \Pr\{x_{S1}|\hat{y}_R, y_{D1}\}^{(0)})$ 
8 find,  $\forall y_R, \hat{y}_R^* = \operatorname{argmin}_{\hat{y}_R} d^{(0)}(\hat{y}_R, y_R)$ 
9  $\Pr\{\hat{y}_R|y_R\}^{(1)} \leftarrow \mathbb{1}_{\hat{y}_R=\hat{y}_R^*}$ 
10 while  $|\Pr\{\hat{y}_R|y_R\}^{(k)} - \Pr\{\hat{y}_R|y_R\}^{(k-1)}| / (|\mathcal{Y}_R| \cdot |\mathcal{Z}|) \geq \epsilon$  do
11    $\Pr\{\hat{y}_R\}^{(k)} \leftarrow \sum_{y_R} \Pr\{y_R\} \Pr\{\hat{y}_R|y_R\}^{(k)}$ 
12    $\Pr\{\hat{y}_R, y_{D1}\}^{(k)} \leftarrow \sum_{y_R} \Pr\{y_R, y_{D1}\} \Pr\{\hat{y}_R|y_R\}^{(k)}$ 
13    $\Pr\{\hat{y}_R|y_{D1}\}^{(k)} \leftarrow \frac{\Pr\{\hat{y}_R, y_{D1}\}^{(k)}}{\Pr\{y_{D1}\}}$ 
14    $\Pr\{x_{S1}|\hat{y}_R, y_{D1}\}^{(k)} \leftarrow \frac{1}{\Pr\{\hat{y}_R, y_{D1}\}^{(k)}} \sum_{y_R} \Pr\{x_{S1}, y_R, y_{D1}\} \Pr\{\hat{y}_R|y_R\}^{(k)}$ 
15    $d^{(k)}(\hat{y}_R, y_R) \leftarrow D_{KL}(\Pr\{x_{S1}|y_R, y_{D1}\} \parallel \Pr\{x_{S1}|\hat{y}_R, y_{D1}\}^{(k)})$ 
16   find,  $\forall y_R, \hat{y}_R^* = \operatorname{argmin}_{\hat{y}_R} d^{(k)}(\hat{y}_R, y_R)$ 
17    $\Pr\{\hat{y}_R|y_R\}^{(k+1)} \leftarrow \mathbb{1}_{\hat{y}_R=\hat{y}_R^*}$ 
18    $k \leftarrow k + 1$ 
19 end
20  $\Pr\{\hat{y}_R|y_R\} \leftarrow \Pr\{\hat{y}_R|y_R\}^{(k)}$ 
21  $r(\beta) \leftarrow \sum_{y_R, \hat{y}_R} \Pr\{\hat{y}_R|y_R\} \Pr\{y_R\} \log_2 \left( \frac{\Pr\{\hat{y}_R|y_R\}}{\Pr\{\hat{y}_R\}} \right) - \sum_{y_{D1}, \hat{y}_R} \Pr\{\hat{y}_R|y_{D1}\} \Pr\{y_{D1}\} \log_2 \left( \frac{\Pr\{\hat{y}_R|y_{D1}\}}{\Pr\{\hat{y}_R\}} \right)$ 
22  $I(r(\beta)) \leftarrow \sum_{x_{S1}, y_{D1}, \hat{y}_R} \Pr\{x_{S1}|\hat{y}_R, y_{D1}\} \Pr\{\hat{y}_R, y_{D1}\} \log_2 \left( \frac{\Pr\{x_{S1}|\hat{y}_R, y_{D1}\}}{\Pr\{x_{S1}|y_{D1}\}} \right)$ 

```

Algorithm 4: Iterative IB algorithm that delivers deterministic quantizer mapping while considering side information, where $\mathbb{1}$ is the indicator function.

G Joint Decoding for Compress and Forward

Similarly as in App. D, a symbol-by-symbol MAP decoder delivers²

$$L(\hat{u}_l) = \log \frac{\Pr\{u_l = 0, \mathbf{y}_{D1}, \hat{\mathbf{y}}_R\}}{\Pr\{u_l = 1, \mathbf{y}_{D1}, \hat{\mathbf{y}}_R\}} \quad (\text{G.23})$$

for the final decision. As these joint distributions are not directly accessible, the set of all possible code words is divided into two subsets $\Gamma_l^{(1)}$ and $\Gamma_l^{(0)}$ containing code words \mathbf{c} whose l th information bit is $u_l = 1$ and $u_l = 0$, respectively.

$$L(\hat{u}_l) = \log \frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \Pr\{\mathbf{c}, \mathbf{y}_{D1}, \hat{\mathbf{y}}_R\}}{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \Pr\{\mathbf{c}, \mathbf{y}_{D1}, \hat{\mathbf{y}}_R\}} \quad (\text{G.24a})$$

$$\begin{aligned} &= \log \frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \Pr\{\mathbf{y}_{D1}, \hat{\mathbf{y}}_R | \mathbf{c}\} \Pr\{\mathbf{c}\}}{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \Pr\{\mathbf{y}_{D1}, \hat{\mathbf{y}}_R | \mathbf{c}\} \Pr\{\mathbf{c}\}} \\ &= \log \frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \Pr\{\mathbf{y}_{D1} | \mathbf{c}\} \Pr\{\hat{\mathbf{y}}_R | \mathbf{c}\} \Pr\{\mathbf{c}\}}{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \Pr\{\mathbf{y}_{D1} | \mathbf{c}\} \Pr\{\hat{\mathbf{y}}_R | \mathbf{c}\} \Pr\{\mathbf{c}\}} \\ &= \log \frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \prod_{i=0}^{n-1} \Pr\{y_{D1_i} | c_i\} \Pr\{\hat{y}_{R_i} | c_i\} \prod_{j=0}^{k-1} \Pr\{u_j\}}{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \prod_{i=0}^{n-1} \Pr\{y_{D1_i} | c_i\} \Pr\{\hat{y}_{R_i} | c_i\} \prod_{j=0}^{k-1} \Pr\{u_j\}} \\ &= \log \frac{\sum_{\mathbf{c} \in \Gamma_l^{(0)}} \prod_{i=0}^{n-1} e^{-(L(y_{D1_i} | c_i) + L(\hat{y}_{R_i} | c_i))c_i} \prod_{j=0}^{k-1} e^{-L(u_j)u_j}}{\sum_{\mathbf{c} \in \Gamma_l^{(1)}} \prod_{i=0}^{n-1} e^{-(L(y_{D1_i} | c_i) + L(\hat{y}_{R_i} | c_i))c_i} \prod_{j=0}^{k-1} e^{-L(u_j)u_j}}. \end{aligned} \quad (\text{G.24b})$$

The first exponential term in (G.24b) represents information about the code bits from the channels ($S \rightarrow D, S \rightarrow R$) and the second exponential term a priori knowledge about the information bits. Both will be given as input to an appropriate decoder like the Bahl Cocke Jelinek Raviv (BCJR) [BCJR74]. It becomes clear that the decoder itself needs not to be modified. Solely, the LLRs $L(\hat{y}_{R_i} | c_i)$ have to be found and added to the LLRs $L(y_{D1_i} | c_i)$ which are delivered by the demapper.

²Indices to distinguish u_1 and u_2 are dropped such that u and c in the following refer to u_1 and c_1 of Fig. 6.6.

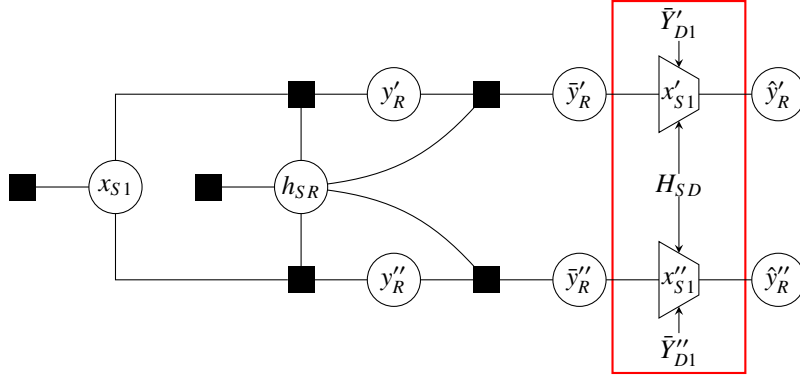


Figure H.3: IBG for explicit equalization using different quantizers on each sub carrier depending on effective SNR.

H Explicit Equalization with Effective SNR and Full Side Information

Theoretically, it is possible to design an individual quantizer for each subcarrier according to its SNR after equalization as shown in Fig. H.3. The drawback, which makes it infeasible in practice, is however that the quantizer rate depends as well on the SNR of the respective subcarriers. Hence, different channel realizations lead to different compression rates with restriction to deterministic quantization. In this case, it is not possible to find a general setup for a practical implementation. Furthermore, time slot parameter τ is directly connected to the compression rate. This problem could be solved by applying general random quantization by fixing the rate for each subcarrier in the design process which requires bisection method explained in Sec. 6.1. Anyhow, the complexity to do this is infeasibly high due to the bisection search. Hence, the scheme with deterministic quantization will be only considered as reference scheme for comparison to the other schemes. The required input pmf for the IB algorithm is

$$\Pr\{x'_{S1}, \bar{y}'_R, \bar{y}'_{D1}, |h_{SD}|\} = \Pr\{\bar{y}'_R | x'_{S1}\} \Pr\{\bar{y}'_{D1} | x'_{S1}, |h_{SD}|\} \Pr\{|h_{SD}|\} \Pr\{x'_{S1}\} \quad (\text{H.25})$$

where pmfs $\Pr\{\bar{y}'_R | x'_{S1}\}$, $\Pr\{\bar{y}'_{D1} | x'_{S1}, |h_{SD}|\}$, and $\Pr\{|h_{SD}|\}$ have to be appropriately sampled from $\mathcal{N}\left(\sqrt{P_{S1}}x'_{S1}, \frac{|h_{SR}|}{2}\right)$, $\mathcal{N}\left(\sqrt{P_{S1}}x'_{S1}, \frac{|h_{SD}|}{2}\right)$, and $\mathcal{R}(d_{SD}^{-\alpha})$. Furthermore, $|h_{SR}|$ is fixed for the quantizer realization in contrary to $|h_{SD}|$ which is a variable in the pmf. In particular, the resulting quantizer holds for a specific effective SNR according to the instantaneous channel $|h_{SR}|$ and not for an ensemble of all possible values of $|h_{SR}|$ which is why $\Pr\{x'_{S1}, \bar{y}'_R, \bar{y}'_{D1}, |h_{SD}|\}$ is not a function of $|h_{SR}|$. The complexity for this pmf is $2^{3 \cdot l_{ad} + \frac{m}{2}}$. Although the complexity of this scheme is less than that of schemes which are valid for the ensemble of channel coefficients h_{SR} , the overall simulation effort is much higher because a lot more quantizers have to be designed due to dependence on SNR per carrier in combination with the SNR of the direct link (side information). The complexity is further increased because source coding rate r depends on the instantaneous

vector channel for an OFDM symbol and thus, has to be averaged over an appropriate number of statistical realizations. Hence, simulation results in Fig. 7.12 are obtained for a coarser resolution compared to the other plots.

